

QY	1870	aLeuGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaVa	1890	Db	4119	CATGACGAGCAGCGGCACCTCCCGAGTGTGTGTACCGCTGTGTACCGGATGGGA	4178
Db	3039	CCTCCAGCAGAGACCCAGTGTCTTACAAACACAGGCATGAAGGGTATCATCACCGCTGT	3098	QY	2250	uGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAl	2270
QY	1890	lGluProSerLysProThrValLeuArgSerThrSerSerProValArgProAl	1910	Db	4179	ACAGACGAGGCCAGCAGATGGGCTCCAAGTCTCCAGGCAACACACGACGCCGCCAGC	4238
Db	3099	GGAGCCAGCAAGCCACGGCTCTGAGGTCCACTCCACTCTCCACCGTTCGCCCCAGC	3158	QY	2270	aPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluI	2290
QY	1910	aAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrPr	1930	Db	4239	CTTCTTCAGCAGCTTCACGAGAGCACTCCGCCATGTCAGTCCAAAGAACAAAGAT	4298
Db	3159	TGCCACATTCACCTGCCACCCACTGCCACTGGCGGCACCTCGATGGGGTCTACCC	3218	QY	2290	eAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGl	2310
QY	1930	oThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluAr	1950	Db	4299	CAACAAGAAGCTGAACACCCACACCGGAATGAGCTGAATACATATATCAGCACCTGG	4358
Db	3219	TACCTCATGAGCCCGCTGTGCTGCCCAAGAGGCCCGCGGTGCGCGGCACAGCG	3278	QY	2310	yThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGl	2330
QY	1950	gProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGl	1970	Db	4359	GACGGAGATCTTCAATATGCGCGCATCACCGGAACAGCGCTTATGACTATAGAGCCA	4418
Db	3279	GCCCCGAGCAGACACCGGCCCATGCTTCTCTCGCAGGCCCCAGCGCTCCGGGTGGA	3338	QY	2330	nAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLe	2350
QY	1970	uProAlaSerSerProSerLysGlySerGluProArgProLeuValProProValSerGl	1990	Db	4419	GGCGGTGCAGGAACATGCCAGCACCAACATGGGGCTGGAGGCCATAATTAGAAAGCACT	4478
Db	3339	GCCGCGCTCTCCCCAGCAGAGGCTCGAGGCCCGGCCCTAGTGCCTCTCTCTG	3398	QY	2350	uMetGlyLysTyrAspGlnTyrGluSerProProLeuSerProProLeuAsnAlaPheAsnPr	2370
QY	1990	yHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisAlaSerProAs	2010	Db	4479	CATGGTAAATATGACCACTGGGAAGAGTCCCGCGCTCAGCGCAATGCTTTTAAACCC	4538
Db	3399	CCAGCCACCATCCCGCCACCCCTGCGAAGAACCTCGACCTCACACGCCAGCCCGGA	3458	QY	2370	oLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSe	2390
QY	2010	pProProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPr	2030	Db	4539	TCTGAATGCCAGTGCAGCGCTGCCCGCTCTATGCCCATACCGCTGTCTGCGGACGGAG	4598
Db	3459	CCCCCGGGGCCACTGCTCGGCTCGAGCCCGCCACCGGGAAAGACTCAAAAGTAAACC	3518	QY	2390	rAspHisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyArgProSe	2410
QY	2030	oPheSerIleGlnLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPr	2050	Db	4599	TGACCACACACTCACCTCGCCAGTGGCGGGGAGGCAAGGTCTCTGGCAGACCCAG	4658
Db	3519	CTTTTCCATCAGGAACCTGGAACCTCCGTTCTCTGGGTATCCACGCGACGACTACAGCC	3578	QY	2410	rSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerVa	2430
QY	2050	oGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLe	2070	Db	4659	CAGCCGAAAGCAAGTCCCGCGCCCGGCTTGGCATCTGGGGACCGGCCACCCCTCTGT	4718
Db	3579	CGAAGGGGTGAGCGCGTACGCTGTGAGCTACCCAGTCTGACCCACGACAGAGGGCT	3638	QY	2430	lSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTr	2450
QY	2070	uProLysHisLeuGluLeuLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGl	2090	Db	4719	CTCTCAGTGCACTCGAGGGAGACTGCAACCGCCGCGACGCGCTCACCACCGCGTGTG	4778
Db	3639	CCCCAAGCACCTGGAAGACTCGACAAGACCACTTGGAGGGGAGCTGCGGCCCAAGCA	3698	QY	2450	pGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetAr	2470
QY	2090	nProGlyProValLysLeuGlyGlyAlaAlaHisLeuProHisLeuArgProLeuPr	2110	Db	4779	GGAGGACAGCCCTCGTCCGAGGTTCACGCAATTCCTTACACCCCTGATCATGCG	4838
Db	3699	GCCAGGCCCCGTGAAGCTTGGCGGGAGGCGCCACCTCCACACTGCGCGCGCTGCC	3758	QY	2470	gLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyPr	2490
QY	2110	oGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGl	2130	Db	4839	GCTGACGCGGGGTGTCATGCTTCCCAACCCCAACCGGGCTCCCGCGGCGACGGGCC	4898
Db	3759	TGAGAGCCAGCCCTCGTCCAGCCGCTGCTCCAGACCGCCCGGAGGGTCAAGGTCAACA	3818	QY	2490	oLeuAlaGlyProHisAlaTrpAspGluGluProLysProLeuLeuLeuCysSerGlnTy	2510
QY	2130	nArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHi	2150	Db	4899	CCTCGCTGGCCCCCACCACCGCTGGGACGAGGAGGCCCAAGCCACTGCTCTGCTCGCAGTA	4958
Db	3819	GCGGGTGTGTCACCTGGCCAGCACATCAGTGAGGTTCATCACAGGACTACACCCGCA	3878	QY	2510	rGluThrLeuSerAspSerGlu	2517
QY	2150	gHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSe	2170	Db	4959	CGAGACACTCTCCGACAGCGAG	4980
Db	3879	CCACCCACAGCAGCTCAGCGCACCCCTGCGCGCCCTCTACTCTTCTTCCCTGGGGCAG	3938	RESULT 7			
QY	2170	rCysProValLeuAspLeuArgProProSerAspLeuTyrLeuProProProAspHi	2190	S83390			
Db	3939	CTGCCCCGCTCTGGACCTCGCGCGCCACCCAGTGACCTTACCTCCCGCCCCCGACCA	3998	LOCUS			
QY	2190	sGlyAlaProAlaArgGlySerProHisSerGluGlyLysArgSerProGluProAs	2210	DEFINITION			
Db	3999	TGGTGCCCCCGCGTGGCTCCCGCCACACGCAAGGGGGCAAGAGGTCTCCAGAGCCAAA	4059	ACCSSION			
QY	2210	nLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGl	2230	VERSION			
Db	4059	CAAGAGCTCGGTCTTGGGTGGTGGAGACCGGTATTGACCTGTGCCACCGGAGGG	4118	KEYWORDS			
QY	2230	yMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGl	2250	SOURCE			
				ORGANISM			
				Hom sapiens (human)			
				Hom sapiens			
				Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
				Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
				REFERENCE			
				1 (bases 1 to 2930)			
				Authors			
				Sande, S. and Privalsky, M.L.			

Db 819 GGTCCAGCTCCAGTCCGCTACTCAGAGCATGCCAAGGCCCGCGTGGGCCCTGTCTACCAT 878
QY 1150 tGlyLeuProLeuProMetAspProLysValLeuAlaProPheSerGlyValLysGlnG 1170
Db 879 GGGGCTGCCCTTCCCATGACCCCAAAAGCTGGCACCTTTCAGCGGAGTGAAGCGGA 938
QY 1170 uGlnLeuSerProArgGlyGlnAlaGlyProGluSerLeuGlyValProThrAlaG 1190
Db 939 GCAGCTGTCCCAACGGGCGCAGGCTGGGCCACCGGAGAGCTGGGGTGGCCACAGCCCA 998
QY 1190 nGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrly 1210
Db 999 GGAGGCGTCCGTGCTCAGAGGGACAGCTCTGGGCTCAGTTCCGGGCGGGAAGCATCACCA 1058
QY 1210 sGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleTh 1230
Db 1059 AGGCATTCACGACACAGGGTGGCCCTCGACAGCGGCATCATACCGGGCTCCATCAC 1118
QY 1230 rHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAs 1250
Db 1119 CCACGGCAGCGCAGCTGACGTCTGTACAAAGGCGCACCATCACCGGGTCCATCATC 1178
QY 1250 pSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTh 1270
Db 1179 CAGCCGAGTCCGTTCGACCGCGCGGAGGACAGCTGCCCAAGGGCCACGTCATCTA 1238
QY 1270 rGluGlyLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSe 1290
Db 1239 CGAAGCAAGAGGGCCAGCTCTTGTCCTATGAGGGTGGCATGCTGTGACCCAGTGCTC 1298
QY 1290 rLysGluAspGlyArgSerSerGlyProProHisGluThrAlaAlaProLysArgTh 1310
Db 1299 CAAGGAGGACGGCAGAAGCAGCTCAGGACCCCGCCCATGAGAGCGCGCCCAAGCGCAC 1358
QY 1310 rTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLe 1330
Db 1359 CTATGACATGATGAGAGGGCGCGTGGGCGAGAGCCATCTCTCAGCGAGCATCGAAGTCT 1418
QY 1330 uMetGlyArgAlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHisH 1350
Db 1419 CATGGCCGTGCCATCCCGCGGAGGCGACACAGCCCGCCACCATCTCAAGAGCAGCACCA 1478
QY 1350 sIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTy 1370
Db 1479 CATCCGCGGTCCATCACACAGGATCCCTCGTCTCTAGTGGAGGCGACAGGAGACTA 1538
QY 1370 rLeuArgArgGluAlaLysLeuLysArgGluGlyThrProProProProProSe 1390
Db 1539 CCTGCGTCGGGAGGCCAAGCTCTAAAGCGGGAGGCGACGCTCCCGCCCGCCCGCCCTC 1598
QY 1390 rArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAl 1410
Db 1599 ACGGACCTTGACCGAGGCTCAAGACGCGAGGCGCTGGGCGCCCTGAGAGCTGAAGCGGCG 1658
QY 1410 aHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProAr 1430
Db 1659 CCATGAGGCGCTGGTGGCCACGCTGAAGGAGGCGCGCTCCATCCATGAGATCCCGCG 1718
QY 1430 gGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluLysE 1450
Db 1719 CGAGGAGCTGCGGCACACGCGCGAGCTGCGCCCTGGCCCGCGCGCTCAAGGAGGGGCTC 1778
QY 1450 rIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLy 1470
Db 1779 CATACGAGGGCACCCCGCTCAAGTACGACCGGGCGGTCCACCATCTGGCTCCAAAAA 1838
QY 1470 sHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLe 1490
Db 1839 GCACGAGCTAGCTCTCCCTCATCGCAGCGCCCGCGGAGGCTTCCACCCCGTGCAACCGCT 1898
QY 1490 uAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSe 1510
Db 1899 GGATGTGATGGCGGCGCGGCACTGGAAACGTGGCTGTCTACGAGGAGAGCGCTGAAGAG 1958

QY 1510 rArgProGlyThrAlaSerSerGlyGlySerIleAlaArgGlyAlaProValIleVa 1530
Db 1959 CCGCCACGAGGACCGCCAGCAGCTCGGGGGCTCCATTGGCGCGCGGCCCGCTCATCTGT 2018
QY 1530 lProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPh 1550
Db 2019 GCGTGAAGCTGGGTAAGCGCGGCGCAGAGCCCCCTGACCTATGAGGACCAACGGGCGACCTT 2078
QY 1550 eAlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuG 1570
Db 2079 TGCCGCGCACCTCCACAGAGTTCCGCGGTGACCATCGGGAGCCACCGCGCGCTGCA 2138
QY 1570 nGluGlySerLeuSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProAr 1590
Db 2139 GGAGGGCAGCTTTCGTCCAGCAGGATCCCGAGGACCGAAAGCTGACGTCGACGCGCTCG 2198
QY 1590 gGluIleAlaLysSerProHisSerThrValProGluHisHisProHisProIleSerPr 1610
Db 2199 TGAGATCGCAAGTCCCGCACAGCAGCCGTGCGGAGCACACCCACACCCCTCTCGCC 2258
QY 1610 oTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLe 1630
Db 2259 CTATGAGCACCTTGTTCGGGGCGTGAGTGGCGTGGACCTGTATCGCAGCCACATCCCT 2318
QY 1630 uAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaIleTyr 1650
Db 2319 GGCTTCGACCCACCTCCATACCCCGCGGATCCCTCTGGAGCGCAGCGCTGCTACTA 2378
QY 1650 rLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuI 1670
Db 2379 CCTGGCCCGACACCTGGCCCGCAACCCACCTACCGGACCTGTACCCACCTACCTCAT 2438
QY 1670 eArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleLeuAsnAspTyrI 1690
Db 2439 CCGGCTTACCCCGCACCGCGCGCTGGAGAACCGGAGACCATCATCAATGACTACAT 2498
QY 1690 eThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLe 1710
Db 2499 CACCTCGCAGCAGATGCACCAACACGCGCCACCGCATGGCCCGCAGCTGATATGCT 2558
QY 1710 uArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgG 1730
Db 2559 GAGGGGCTCTCGCCCGCGAGTCTCTGCTGGCACTCAACTACGCTGCGGGTCCCGGAG 2618
QY 1730 yIleIleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyTh 1750
Db 2619 CATCATCGACCTGTCCCAAGTGCACACCTGCTGTGCTCGTGCCTCCCGCCGACACGAG 2678
QY 1750 rProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSe 1770
Db 2679 CCCAGCCACCGCATGGACCGCTTGCCTACCTCCCGCCCGCGCCCGCTTTCAGCAG 2738
QY 1770 rArgHisSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrTh 1790
Db 2739 CCGCCACAGCAGCTCCCACTCTCCCGAGGAGTCCACACACTTGACAAAACCAACAC 2798
QY 1790 rThrSerSerGluArgGluArgAspArgGluArgAspArgAspArgGluAr 1810
Db 2799 CAGCTCTCGTCGAGCGGAGCGAGACCGGATCGAGAGCGGACCGGATCGGAGCG 2858
QY 1810 gGluLysSerIleLeuThrSerThrThrValGluHisAlaProIleTrpArgProGl 1830
Db 2859 GGAAAGTCCATCTCCCTCCACCGAGCTGGAGCAGCAGCCATCTCTGGAGACCTGG 2918
QY 1830 yThrGluGlnSerSerGlySerSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1850
Db 2919 TACAGAGCAGAGCAGCGCAGCAGCGCAGCGCGGGGGTGGGGGCGAGCAGCGCG 2978
QY 1850 qProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAl 1870
Db 2979 CCGCGCTCCCACTCCCATGCGCCACGACCTCGCCCATCTCCCTTCGGACCCGAGATGC 3038

6857 TGTCCTCAGTACACTCAGAGGGGACTGCAATGCGGACACCACTCACCAACCGTGT 6916

2449 lTtpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMe 2469

6917 GTGGAGGACCGGCCCTCATCTCGAGGGTCCAGCCATTCCTCCACACCCCTTGATTAT 6976

2469 tArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerG1 2489

6977 GAGGTACAGGAGGTGTCTATGCGCTCCCGCCGCCACCTGGCCCTTGGCGGAGGAGCGG 7036

2489 yProLeuAlaGlyProHisHisAlaTtpAspGluGluProLyseProLeuLeuCySerG1 2509

7037 GCCCTAGCTGTGCTCCACACCGCTGGATGAGGAGCCCAAGCCACTGCTGTGTTTACA 7096

2509 nTyrGluThrLeuSerAspSerGlu 2517

7097 GTATGAGACACTCTCGGACAGCGAG 7121

RESULT 6

HSU37146 5989 bp mRNA linear PRI 31-OCT-1995

LOCUS Human silencing mediator of retinoid and thyroid hormone action

DEFINITION (SWRT) mRNA, complete cds.

U37146

U37146.1 GI:1045654

Accession

Version

Keywords

Source Homo sapiens (human)

Organism Homo sapiens

Reference 1 (bases 1 to 5989)

Authors Chen, J.D. and Evans, R.M.

Title A transcriptional co-repressor that interacts with nuclear hormone receptors

Journal Nature 377 (6548), 454-457 (1995)

Medline 96008552

PubMed 7566127

Reference 2 (bases 1 to 5989)

Authors Chen, J.D. and Evans, R.M.

Title Direct Submission

Journal Submitted (27-SEP-1995) J. Don Chen, Gene Expression Lab, The Salk Institute, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Features

Location/Qualifiers

1..5989

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/cell_line="HeLa"

1..5989

/gene="SMRT"

496..4983

/gene="SMRT"

/note="transcriptional co-repressor"

/codon_start=1

/product="silencing mediator of retinoid and thyroid hormone action"

/protein_id="AAC50236.1"

/db_xref="GI:1045655"

/translation="MEAWDHPDKFAFAEAQKLPDGPCCWTSGLPPPPPREVIKAS PHADPFAFYPFPHLPGLHDTARPVLPFRPTISNPPPLISSAKHPSVLERQIGA ISQMSVOLHPVSEHAKAPVPGVTMGLPMDPKLAPFSGVKQEQLSPRGAGPPE SLGVTAGAESVLRGTALGSPVGGSTKGIPTSTRVPSDSAITVRSITHTGTPADLYK GTITRIIGEDSPRLDGRDLSLPKHVIVYEGKGVHLSYEGGMSVTQCSKEIDGRSS GPETHAAPKRTYDMGSRVGRALSSAIEGLMGRAPRPHSHPHLKEOHILGRSIT QGIPRSVYAEQDYLRLREKULKEGTPPPPPSRDTEAYKTQALGLKLKPAHEGL VATKEAGRSITHEPRELRTPELPAPRLEKGSITQGPLKIDYTGASTTSGKGD VRLSGSGRTFPVPHLDVNMADARALERCYSLESLKSPGTASSSGSGSTARAPVIV PELGKPROSLTYEDHGAFFAGHLPGRSPVTMRPELPLOEGLSSSSKASQDRKLTST PREAKSPHSVTPRHHPHPI SPYEHLLRGVSDLYRSHIPLAFDPTSIPRGIPLDAA AAYLPHRLANPNYTPHLYPPYLIRGVYDPALENROTIINDYITSOQMHNTATAMA QADMLRGLSPRESSLALNYAAGPRGIIDLSOVPHLPVLVPTPTGATAMDRLYLP TAPQFSSRHSSTPLSPGPGTHLTPITTSSESRDRDRDRDREREKSLTSTTT VEHAPIWRPGTEQSSGSGSGSGSSSRPASHAHQHSPISPRTDALQORPSVL

HNMGKGIITAVEPSKPTVLRSTSTSPVRPAATFPPTHATHCPLGGLTLDGVVPTLMEPV
LLPKAPVAVPERPRADTGHAFKADPPARGSLPASPSPSGSEPRPLVPVSGHATI
ARTPAKLIAPHAHPDPAPASADPHREKTQSPKPSIOELRLSLGHSYSPGEG
VEFVSPVSSPSLTHDKGLKHELDKSHLSELRPKQPGPVKLGSAALHLPHLPSFG
ESQSPFLUQTAPGCVKGRVVTLAQHSIVITQDYTRHHFQQLSAPLAPLPSFGP
ASCPLDURPPSDLYLPDPHGPARGSPHSGGKRSPEPKTSVLGGEDGIDFVS
PPGEMTEPHRSASVYPLLYRDGEQTEPSRMGSKSPGNTSQPPAFSKLTESNAWK
SKKQINKKLANTHNRNEPYNISQGTETFNMPATGTCLMTYRSQAOEHASTNGL
LIRKALMGKYDOWEESPPLSANAFNPLNASLPAAMPTAAGDRSDHGLTSPGGG
GKAKYSPSRKAKSPAPGLASGDRPPSVSVHSEGDNRRTPLTNRWEDRPSAG
STPPYFNPLIWRKQVNASPPPPGLPAGSGPLAGPHANDEEPKLLCSQVETILSDS
E"
4483..4620
/gene="SMRT"
/note="alternatively spliced insert"

exon

ORIGIN

Alignment Scores:

Pred. No.: 9,78e-152 Length: 5989

Score: 7885.00 Matches: 1510

Percent Similarity: 94.34% Conservative: 7

Best Local Similarity: 93.91% Mismatches: 14

Query Match: 59.67% Indels: 77

DB: 9 Gaps: 7

US-09-522-753-5 (1-2517) x HSU37146 (1-5989)

984 ProProGluAspAlaAlaProThrLysProAlaProProAla----- 998

168 CCACCAAGA-----TCACCATATGAGCCACGACCCCTGCATAGGAATAATCATAT 218

999 ProProProPro----- 1002

219 CCGAGCCACCTGTGTTTCATAAGCCATCTGCCATTCGTATGCGGTCCATGTATCTATT 278

1003 -----GlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGly 1016

279 GCAGAGTCCCAAGTTTTCATCAGCACTGAAACCAACCATCCGTCGTCGGAGAGGAGGAGTGT 338

1017 -----SerSerPro-----Arg 1020

339 CTCCTCTCTCTGCATAGGCATGAGGTCTCCCTCTAGTGTGTGTGTGGAGGAGAGA 398

1021 GlyLysSerArgSer-----ProAlaProProAla----- 1030

399 GGAAGATTCCGAGCTCTGTACCAACCCCGCCCTCGTCGGAGAGGAGGAGTGT 458

1030 ----- 1030

459 CTTGACGAGGGCTCATATCATATATCTTCTAGATGGAGGATGGAGCCGCCACCC 518

1031 -AspLysGluAlaPheAlaAlaGluAlaGlnLysLeuPrcGlyAspProCysTrpTh 1050

519 CGAAGAGAGGCTTCGAGCCGAGGCCAGAGCTGCTCGGAGACCCCTTGTCTGGAC 578

1050 rSerGlyLeuProPheProValProArgGluValLileLysAlaSerProHisAlaPr 1070

579 TTCGGCCTGCTCTCCCGTGCCTCCCTGAGGTGATCAGAGCTCCCGCATGCC 638

1070 oAspProSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAs 1090

639 GGAACCTCTAGCTCTCTTACCTCCCTGAGTCCACCTGCTCCCTTCCCTTCCATGA 698

1090 pThrAlaArgProValLeuProArgProProThrLileSerAsnProProProLeuLileSe 1110

699 CACTGCCCGCCGCTCTCGCGCCGCCACCCACCATCTCCAAACCCGCTCCCTCATCTC 758

1110 rSerAlaLysHisProSerValLeuGluArgGlnLileGlyAlaLileSerGlnGlyMetSe 1130

759 CTCTGCCAAGACACCCAGCGCTCTCGAGAGGCAATAGTGCATCTCCCAAGAAATGTC 818

1130 rValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMe 1150

1714 rProArgGluSerSerLeuAlaLeuAenTyrAlaAlaGlyProArgGlyIleIleAspLe 1734
Db ACCGCGAGAGTCTCGCTGCGCCCTCAATATATGCGCGTGGCCCAAGAGGCATTATCGACCT 4819
1734 uSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAl 1754
Db GTCCCAAGTGGCCACACTCGCGCGTGTGTGGCCACCACCAAGCCAGGACCCCTGCGCACCGC 4879
1754 aMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSe 1774
Db CATCGACCGCTTCCCTACTCCCTCACTGGCCCCCACCCTTCAGCAGCGGCCACAGTAG 4939
1774 rSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSe 1794
Db CTCAACGCTGTCCAGGAGGCCCACTCACTAGCTAAACCAACTGCCACATCTTCATC 4999
1794 rGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSeril 1814
Db GGAGCGGGAACGGGAACGTGAGCGGGAACGAGAC-----AAGTCCAT 5041
1814 eLeuThrSerThrThrThrValGluHisAlaProIleTTrpArgProGlyThrGluGlnSe 1834
Db CCTCAGCTTACCACTACAGTGAGAGATGACCCATCTGGAGACCTGTGTACGGAGCAGAG 5101
1834 rSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAlaSerHi 1854
Db CAGCGGGCT-----GGGGGCGAGCAGCGCGCGCGCTCCCA 5137
1854 sSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnGlnAr 1874
Db CACC-----CACCAGCACTCGCCCATCTCCCCCGGACCCAGAGCGCTTGCAGCAGAG 5191
1874 gProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLy 1894
Db GCCAGTGTGTGCACACACGAGCATGAAGGCGTGGTGCACCTCGTGGAAACCCCGCAC 5251
1894 sProThrValLeu-----ArgSerThrSerThrSerSerProValArgProAlaAl 1911
Db GCCACCGTCTCGAGTGGGCCAGGTCCACCTCCACTCTTCGCTGTCCGCGCAGCTGC 5311
1911 aThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProTh 1931
Db CACATTCCCACTGCCACCACTGCCCACTTGTGTGGACCCCTTGAAGGGGTCTTACCCCTAC 5371
1931 rLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgPr 1951
Db CCTCATGGAGCCGCTCTGTATCCCAAGGAGACTCTCGGGTCCGCGCGCGAGCGGC 5431
1951 oArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluPr 1971
Db CCGGGTGGACGCTGGCCATGCTTCTTACCAAAACCCCGGGCGG-----GAGCC 5482
1971 oAlaSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHi 1991
Db CGCCTCTCACCAGCAAGAGCTCCAGGCGCCGATCCCTAGCACCCCGCGCTCCAGCCA 5542
1991 sAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspPr 2011
Db CACAGCATCGCCCGCACCCACCAAGAACCTTGCACCCACCATGCCAGTCCGAGCCC 5602
2011 oProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPh 2031
Db GCCCGCG-----CCCACTCGGCTCAGATCTGCACCGGAGAAAGACTCAAGTAAACCTT 5659
2031 eSerIleGlnGluLeuLeuArgSerLeuGlyTyrHis--GlySerSerTyrSerPr 2050
Db TTCATCCAGAAATTGAATCCGTCTCTGGGTATACACAGTGGAGCTGGCTACAGCCC 5719
2050 oGluGlyValGluProValSerProValSerProSerLeuThrHisAspLysGlyLe 2070
Db CGATGGGGTGGAGCCCATCAGCCCGGTGAGTCTCCCCAGCTGACCCACGACAAAGGGCT 5779
2070 uProLysHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysG 2090

5780 CTCCAAACCTCTGGAGAGCTAGAGAAGAGCCACTTGGAAAGGGAGCTCGCGCACAGCA 5839
2090 nProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuPr 2110
Db GCAGCGCCCATGAAGCTCAGCGCGAGGCTGCCCATCTCCACCATCTCGCGGCCTGCG 5899
2110 oGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisG 2130
Db CGAGAGCCAGCCCTCATCCAGCCCATCTCTCAGACTGCCCCAGGATCAAAAGGTCA 5959
2130 nArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHi 2150
Db GAGGGTGGTCACTCGCTCAGCATCAGCAGAGGTCAATACGAGACTACACGCGCCA 6019
2150 sHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSe 2170
Db CCACCCGAGCAGCTCAGTGGCCCCCTTCCCGCCCTCTCTACTCTCTTCCCGGAGCCAG 6079
2170 rCysProValLeuAspArgProProSerAspLeuTyrLeuProProProAspHi 2190
Db CTGCCCTGTCTGGATCTTCGCGCGCCACCCAGTGACCTCTACCTCCACCCCGCACCA 6139
2190 sGlyAlaProAlaArgGlySerProHisSerGluGlyLysArgSerProGluProAs 2210
Db TGGCACCCCGAGCCCGGGATCCCCACAGTAGTGAAGGGGCAAAAGGTCCCCAGAACCCAG 6199
2210 nLysThrSerValLeuGlyGlyGluAspGlyIleGluProValSerProProGluG 2230
Db CAAAACATCGTCTCGGCGAGCAGGAGTGCATTGAGCCCTGTCTCCACAGAGGG 6259
2230 yMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyG 2250
Db CATGACTGAGCCAGGACATGCTCGGAGCACTCGTACCCACTGCTGTATCGAGACGGGA 6319
2250 uGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProAl 2270
Db ACAGGCGGAGCCCC---AGGATGGGTCTAGAGTCTCCAGGCAACACAGCAGCCCGCAAC 6376
2270 aPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGlu 2290
Db CTTCTTCAGTAGCTGACTGAGAGCACTCCGCGCATGGTGAAGTCCAGAGCAGAGAT 6436
2290 eAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProG 2310
Db CAACAAGAAACTCAACACCCACAACCGGAACGAGCCAGAAATACAATATTGGCCAGCCTGG 6496
2310 yThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerG 2330
Db GACGGAAATCTTCAACATCGCCCATCTCGAGCAGCCCTTATGACCTGTAGAAGCCA 6556
2330 nAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleArgLysAlaLe 2350
Db GCGGTGCAAGAAACACCCAGCACCACATGGGGCTAGAGCCATTATTAGAAGGCACCT 6616
2350 uMetGlyLysTyrAspGlnTrpGluSerProProLeuSerAlaAsnAlaPheAsnPr 2370
Db CATGGTAAATATGATCAGTGGGAAGAGCCCCCGCGCTCGGCGCCAATGCTTTTAAACCC 6676
2370 oLeuAsnAlaSerAlaSerLeuPro---AlaAlaMetProIleThrAlaAlaAspGlyAr 2389
Db TCTGAATGCCAGCGCCAGTCTCGCGCTGCTGATGCCCATTAACCACTGCTGACCGAGC 6736
2389 gSerAspHisThrLeuThrSerProGlyGlyLysAlaLysValSerGlyArgPr 2409
Db GAGTGACACGCACTCACTCGCCAGTGGAGGTGGAAAGCCAAAGTCTCTGCGCAGACC 6796
2409 oSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProSe 2429
Db TAGCACCCGAAAGCCCAAGTCGCAGCACACAGGCTTAGCGTCCGGAGACCGACCCCTTC 6856
2429 rValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgVa 2449

Db 2660 GGACAGTACCCCAAGCCAGTTCCTCCCTGTGCTCCACCCAGCAGCACCTACAGCC 2719
Qy 1007 oGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAl 1027
Db 2720 AGAGGGTGACGTCTCTCAGCAGTCGGGAGAGTCCACGTGGCAAGTCCTCCGAGCCAGT 2779
Qy 1027 aProProAlaAspLysGlu-----AlaPheAlaAlaGluAl 1039
Db 2780 GCCTCTCCGCGAGAAAGAGCAGAGAACCCCGCATCTTCTCCGGCTTTTCCCACTGAGGG 2839
Qy 1039 aGlnLysLeuProGlyAspProCysTrpThrSerGlyLeuProPheProValProPr 1059
Db 2840 CCCAAGCTACCGACTGAGCCCCCAGCCAGCTGTCTATCGGGCTGCCCTTCCCATCTCC 2899
Qy 1059 oArgGluValLysAlaSerProHisAlaProAspProSerAlaPheSerThrAlaPr 1079
Db 2900 ACGGGAGGTATCAAGACTTCCCAACACACGCGCTGACCCCTCTGCTCTCTCTACACAC 2959
Qy 1079 oProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgPr 1099
Db 2960 CCCCCGTACCCGCTGCTCTGGGCTTCCACGATAGTGGCCCGCCGCTCTGCTGCACGTC 3019
Qy 1099 oProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuGl 1119
Db 3020 CCCC---ATCTCTAAACCCCCACCCCTCATCTCTCTGCGCAAGCATCCCGCGTACTTGA 3076
Qy 1119 uArgGlnIleGlyAlaIleSer---GlnGlyMetSerValGlnLeuHisValProTyrSe 1138
Db 3077 GAGGACGCTGGTGCCATCTCCACGACGAGGGATGTCACTCCAGTCTGCTGTGCTCACTC 3136
Qy 1138 rGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspPr 1158
Db 3137 AGCATGTCACGAGCCCGCCCGGCTCTCACCATGGGCTGCCCTTCCCGTGGACCC 3196
Qy 1158 oLysLysLeuAlaProPheSerGlyValLysGlnGlnLeuSerProArgGlyGlnAl 1178
Db 3197 TAAGAAGCTG----- 3206
Qy 1178 aGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyTh 1198
Db 3207 -----GGGAC 3211
Qy 1198 rAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValPr 1218
Db 3212 AGCATGGCTCCCGCCACCACTGGAAGCATCACCAAGGCTCCCGAGTACCCGGGCTGC 3271
Qy 1218 oSerAspSerAlaIleThrThrArgGlySerIleThrHisGlyThrProAlaAspValLe 1238
Db 3272 AGACGGCCCC-----AGCTACAGAGGCTCTATCACCCACGCGCACGCCGCGTCTCT 3325
Qy 1238 uTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGl 1258
Db 3326 CTACAAGGGTACCATCAGCAGATCGTCGGTGAGACAGCCCAAGTGCCTTGACCCGGC 3385
Qy 1258 yArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysGlyHisValLe 1278
Db 3396 ACAGAGACACCCCTGCCCAAGGCCATGTCTATGAGGGCAAGAAAGCCACGTCCT 3445
Qy 1278 uSerTyrGluGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSe 1298
Db 3446 ATCTATGAAGTGTGTATGTCGTGTATGTCAGTGTCTTAAGGAGGATGGAAGGACGCTC 3505
Qy 1298 rGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgVa 1318
Db 3506 GGGCCACCCCATGATGATGCGGCCCTTAACGACCACTATGATGATGAGGGCCCGTGT 3565
Qy 1318 lGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGl 1338
Db 3566 AGGAGGACTGTCACTTCAGCCAGCATAGAGGACTCATGGCCCGCGCATC---CCTGA 3622
Qy 1338 uArgHisSerProHisLysLeuLysGluGlnHisIleArgGlySerIleThrGlnGl 1358
Db 3623 GCAGCACAGCCCC---CACCTCAAGGAGCAGCATCACATCCGAGGCTCCATCAGCAGAG 3679

Qy 1358 yIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLe 1378
Db 3680 CATCCCGAGGTCTATGTGGAGCGCAGAGGACTACTTACGGCGGAGGCCCAAGCTCTT 3739
Qy 1378 uLysArgGluGlyThrProProProProProSerArgAspLeuThrGluAlaTyrIly 1398
Db 3740 GAAGCGAGAGGACACACACACCCACCCACCTCGGACCTGACTGAGACTACAA 3799
Qy 1398 sThrGln-----AlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuVa 1415
Db 3800 GCCCGGCGCCCTGGACCTCTGGGTCCCTGAAGCTGAAGCCGACTCAGAGGGTGTGGT 3859
Qy 1415 lAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHi 1435
Db 3860 AGCAACTGTGAAGGAGGCGGCGCTCTATCCATGAGATCCCGAGAGGAGCTGCGCG 3919
Qy 1435 sThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyTh 1455
Db 3920 CACACTGAGCTTACCCCTGGCACACCGGCTCTGAAGGAGGGTTCCATCACCCAGGGCAC 3979
Qy 1455 rProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSe 1475
Db 3980 CCACTCAAGTACGACTCTGGGGCACCTTCCCTGACCCGCTGGACATAATGGCTGA 4039
Qy 1475 rLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAs 1495
Db 4040 CATCATCGGACGCGCGCGCTTTCCTGCCCTGCACCCGCTGGACATAATGGCTGA 4099
Qy 1495 pAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLysLysSerArgProGlyThrAl 1515
Db 4100 CGCCCGGACACTGGAGCGTGCCTGTATGAAGAGAGTCTGAAGAGCGGCTCAGGAGACAG 4159
Qy 1515 aserSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLy 1535
Db 4160 CAGTGTGTGAGGGGGCTTCCATCACGTGGGCTCCAGTCTGCTGCTGCTGAACTGGGCA 4219
Qy 1535 sProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuPr 1555
Db 4220 GCCACGCAAGGCGCTTACGAAGACACACCGGCGCACCTTCCACGACTCACCTGGCC 4279
Qy 1555 oArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlySerLeuSe 1575
Db 4280 ACCTGCTCCCTGTGACACGAGGAGCGCCAGCCACCGCTTCAGAAAGGACGCTCCT 4339
Qy 1575 rSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSe 1595
Db 4340 ATCCACGAAGGCTCCGAGACCGGAAGCTGACATCTACACCCCGGAGATCGCCAAGTC 4399
Qy 1595 rProHisSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLe 1615
Db 4400 CCCACACAGCACTGTGCGGAGCACCCCTTCCACCCCTTCCCTTATGAGCACTTGCT 4459
Qy 1615 uArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProTh 1635
Db 4460 CCGGGCGGTGACGGGTGGACCTGTACCTGTGTCAATCCCATTTGGCTTTCACCCAC 4519
Qy 1635 rSerIleProArgGlyIleProLeuAsp---AlaAlaAlaAlaTyrIleuProArgHi 1654
Db 4520 CTCCATACCCCGAGGATCCCTCTGGAAGCAGCAGCGCGAGCTACTACCTGCGCGGCA 4579
Qy 1654 sLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrIleArgGlyTyrPr 1674
Db 4580 CTTTGGCCCGGACCCACCTTACCCACCTGTATCCACCTTACTCTATCCGCGGCTACCC 4639
Qy 1674 oAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGl 1694
Db 4640 TGACACGGCGGCGCTGGAGAACCGCCAGCACCATCATCATGACTACACTCCGCGAGCA 4699
Qy 1694 nMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSe 1714
Db 4700 GATGCACCAACGCTGCTCCCTCCGCCATGGCCAGCGTGTGACATGTGTGAGGGGTCTGTC 4759

[illegible]

1614 GTGGGCTCCAAGACCGTGTCCCAAGTGTAAAGAACTTCTACTTCAACTCAAGAAGAGCGAG 1673

661 AnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArg 680

1674 AACCTGGACGAAATCTCTTACGACGACACAGCTAAAGATGTAGAAGGAGGACGCTCGG 1733

681 ArgLysLysLysAlaProAlaAlaSerGluGluAlaAAlaPheProProValVal 700

1734 AGCAAGAAGAAGAAGACCCAGCTGTGGGGGAGCGAGGAGACAGACCTTCCCACTTCGCGCT 1793

701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetValGluGlu 720

1794 GAGGACGAAGAGATGGAAGCATCAGGCGCAAGTGCCTAAATGAGAAGAGCTGGCGAGGAG 1853

721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArg--GlyGluCysSerGlyPro 739

1854 GCAGAAGCCTCACAGGCTCTGGGAATGAGTTCCAGAGTTGGGAGTGCAGTGGCCCA 1913

740 AlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAla 759

1914 GCTGTGTCAACAACAGCTGTGATCTGAGAGTGTCCATCCCGCGTTTCAAGACCACG 1973

760 LysAspThrGlyGlnAsnGlyProLysProProAlaThr-----LeuGlyAlaAspGly 777

1974 AAGGACACT-----GGGCTTAACCCACTGGCAGTGAAGCAATCCCGCTGCCACC 2024

778 ProProGlyProProThrProProArgArgThrSerArgAlaProIleGluProThr 797

2025 CAGCCACTGTCTCT-----CTTCAGAAGAACCGGAGTAGCCCTCTGTAGCCCTCC 2078

798 ProAlaSerGluAlaThrGlyAlaProThrProProAlaProProSerProSerAla 817

2079 CCAGTCCCTGATGCAGCTGGCGCCACCATCCCGAGAGCTTC-----CCATCACTGCCGCA 2134

818 ProProProValValProLysGluGluLysGluGluThrAlaAlaAlaProProVal 837

2135 CCCCCTGCTACTGTGGCAAGGATGAACAAGAGCCCGGCTGCTCCAGCTCCCCAGACA 2194

838 GluGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaAspThrGlyLys 857

2195 GAGGATGCCAGGAGCAGAGTCTGAGCGCCGAGG-----ATCGATGTGGAAAA 2245

858 Ala-GluGluPro-----ValLysSerGlyCysThrGlu 868

2246 GCCAGAGAGCCCGAGGCTTCTGAGAAGCCCGCCAGAGTGTAAAGAGTGAACCAAGAA 2305

868 uGluAlaGluGluProAla--LysGlyLysAspAlaGluAlaAlaGluAlaThrAl 887

2306 GGAGACCCGAGAAGAGCTTGAAGCAAAAGCCAAAGGCGCAGAGGCCATTGAAACTGTGTC 2365

887 adGlyAlaLeuLysAlaGluLysLysGluGlySerGlyArgAlaThrThrAlaLys 907

2366 TGAGGCACCATTAAGGTGGAG--TAGGCTGTGTACCAAGGAGCTGTGACCAAGGGTTC 2422

907 sSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGluValAs 927

2423 CAGCTCAGGTGCCACCCAGGACAGTGCATCCAGTGCACCTCAGTGCATGAGGTGGA 2482

927 pGluAlaGluGlyAspLysAsnArgLeuLeuSerProArgProSerLeuLeuThrPr 947

2483 CGAAGCCCGAAGGAGGTGCAAGGGCAGGCTGTCTGCCAACAGGCCACGCTCTCTCACCCC 2542

947 oThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLys 967

2543 GCTTGAGATCTCCCGGCGCGTGTACTCTGCCCCCAAGAGCCGCTTGAGCTTGAAGAGCTGAA 2602

967 sGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProProArgGlu 987

2603 GCAGCGAGCAGCCGCCATCCCCCTATC---GTCAACAAGGTCCATGATGACCCCCCGGGA 2659

987 uAspAlaAlaProThrLysProAlaProProAlaProProProGlnAsnLeuGlnPr 1007

Db 3397 ACACCCCCGGTCACCGCTGCTCTGGCCCTCCACGATAGTCCCGCGCCGCTCTCCGCCA 3456
Qy 1098 ArgProThrIleSerAenProProLeuIleSerSerAlaIshisProSerVal 1117
Db 3457 CGTCCCCCCC---ATCTCTAACCCCCCACCCTCATCTCTCTCCCAAGCATCCCGCGTA 3513
Qy 1118 LeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyr 1137
Db 3514 CTTGAGAGGAGCTGGGTGCCATCTCCAGGGATGTCACTCAGCTTCGTGTGCTCAC 3573
Qy 1138 SerGluHisAlaIshisAlaProValGlyProValThrMetGlyLeuProLeuProMetAsp 1157
Db 3574 TCAGAGCATGCCAAG---CCCATGGGCCCTCTCACCATGAGCTGCCCCCTTGGCGTGGAC 3630
Qy 1158 ProIshisLeuAlaProPheSerGlyValIshisGlnLeuSerProArgGlyGln 1177
Db 3631 CCTAAGAAGCTG----- 3642
Qy 1178 AlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGly 1197
Db 3643 -----GGG 3645
Qy 1198 ThrAlaLeuGlySerValProGlyGlySerIleThrIshisGlyIleProSerThrArgVal 1217
Db 3646 ACAGCACT-GGCTCCGCCACCAAGTGGAGCATCACCAGGGCCT-CCCAGTACCCGGCT 3703
Qy 1218 ProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspVal 1237
Db 3704 GCAGACGGCCCC-----AGCTACAGAGGCTCTATACCCACGC-ACGCCCGCAGACGTC 3756
Qy 1238 LeuTyrIshisGlyThrIleThrArgIleGlyLeuAspSerProSerArgLeuAspArg 1257
Db 3757 CTCTACAAAGGTACCATCAGCAGATCGTGGTGAGGACAGCCCAAGTCGCTTGACCG 3816
Qy 1258 GlyArgGluAspSerLeuProIshisGlyHisValIleThrGluGlyIshisVal 1277
Db 3817 GCACGAGAGGACACCTTGGCCCAAGGGCCATGTCTATCAGGGCAGAAAGGCCACGTC 3876
Qy 1278 LeuSerTyrGluGlyMetSerValThrGlnCysSerIshisGlyLeuAspGlyArgSer 1297
Db 3877 CTATCTTAAAGGTGTATGTCCGTGTACAGTCTTAAGGAGGATGAAGAGGAGC 3936
Qy 1298 SerGlyProProHisGluThrAlaAlaProIshisArgThrTyrAspMetGluGlyArg 1317
Db 3937 TCGGGCCCCACCCATGAGCTGCGGCCCTTAACGCACCTATGACATGATGGAGGCGGT 3996
Qy 1318 ValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPro 1337
Db 3997 GTAGGCAGGACTGTCACTCAGCCAGCATAGAGGACTCATGGCCCGCGCATC---CCT 4053
Qy 1338 GluArgHisSerProHisGlyHisLeuGlyGlnHisIleArgGlySerIleThrGln 1357
Db 4054 GAGCAGCACAGCCCC---CACTTCAAGAGGAGCATCACATCCGAGGCTCCATCAGCCAA 4110
Qy 1358 GlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaIshisLeu 1377
Db 4111 GGATCCCGAGGTCTATGTGAGGCGCAGAGGACTTACGGCGGGAGGCCAAGCTC 4170
Qy 1378 LeuIshisArgGluGlyThrProProProProProSerArgAspLeuThrGluAlaTyr 1397
Db 4171 TTGAAGCGAGAAGGAGACACCACTTCCGCCACCACTCTGGGACCTCGGACCTGAGACCTAC 4230
Qy 1398 LysThrGln-----AlaLeuGlyProLeuIshisLeuIshisGluGlyLeu 1414
Db 4231 AAGCCCCGGCCCTGGACCTCTGGGTCCCTGAGCTGAAGCGCACTCAGAGGGTGTG 4290
Qy 1415 ValAlaThrValIshisGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArg 1434
Db 4291 GTAGCAACTGTGAAGGAGGGCGGCTCTATCCATGAGATCCGAGAGGAGTGCCTG 4350
Qy 1435 HisThrProGluLeuProLeuAlaProArgProLeuIshisGlySerIleThrGlnGly 1454
Db 4351 CGCACACCTGAGCTACCCCTGGCACCCACCGGCTCTGAAGAGGGGTTCATCACCCAGGGC 4410

Qy 1455 ThrProLeuIshisTyrAspThrGlyAlaSerThrThrGlySerIshisHisValArg 1474
Db 4411 ACCCACTCAAGTACGACTCTGGGCACCTCCACTGGCACCAAGAAACACGACGTGGC 4470
Qy 1475 SerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAla 1494
Db 4471 TCCATCTGGGACCGCCCGCGGCTTTCCCTCTGCCCTGCACCGCTGACATAATGGCT 4530
Qy 1495 AspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuIshisSerArgProGlyThr 1514
Db 4531 GACGCCCGGGCACTGAGCGTGCCTCTATGAAGAGAGTCTGAAGAGCGGTCAGGAC 4590
Qy 1515 AlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGly 1534
Db 4591 AGCAGTGGTGCAGGGGCTCCATCACACGTGGGCTCCAGTCTGCTGCTGAACCTGGC 4650
Qy 1535 LysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeu 1554
Db 4651 AAGCCACCGCAAGCCCACTGACTTACGAAGACACCGGGGCACTTACCACTCACCTG 4710
Qy 1555 ProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeu 1574
Db 4711 CCAGTGGCTTCCCTGTGACCAAGGAGGAGCCACCGCCTTCCAGGAAGCAGCCTC 4770
Qy 1575 SerSerSerIshisAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLys 1594
Db 4771 CTATCCAGCAAGGGCTCCCGAGGACCGGAAGCTGACATCTACACCCCGGAGATCGCAAG 4830
Qy 1595 SerProHisSerThrValProGluHisHisProHisProIshisSerProTyrGluHisLeu 1614
Db 4831 TCCCCACACAGCACTGTGCCGAGCACCACCTTCCACCCCTTCCCTTATGAGCACTTG 4890
Qy 1615 LeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspPro 1634
Db 4891 CTCGGGGGCTGACTGGTGTGACCTGTACCGTGTACATCCATTCCTTGGCTTTGACCCC 4950
Qy 1635 ThrSerIleProArgGlyIleProLeuAsp---AlaAlaAlaIshisTyrIshisLeuProArg 1653
Db 4951 ACCTCCATCCCGAGGGATCCCTCTGGAAGCAGCAGCGCCAGCTTACCTGCGCCGG 5010
Qy 1654 HisLeuAlaProAsnProThrTyrProHisLeuTyrProTyrIshisLeuIleArgGlyTyr 1673
Db 5011 CACTTGGCCCCAGCCCACTACCCACACCTGTACCCACCTTACCTCATCCCGCGCTAC 5070
Qy 1674 ProAspThrAlaAlaLeuGluAsnArgGlnThrIleIshisAspTyrIleThrSerGln 1693
Db 5071 CCTGACAGCGCGCCCTGGAGAACCCGAGACCATCATCATGACTACATCATCTGCAG 5130
Qy 1694 GlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeu 1713
Db 5131 CAGATGACCAACAACGCTGCTCCGCTGCGCCAGCTGCTGACATGCTGAGGGGTCTG 5190
Qy 1714 SerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAsp 1733
Db 5191 TCACCGCGAGGTCTCTCGTGGCCCTCAATTAATTCCTGCTGGCCCAAGAGCATTTACGAC 5250
Qy 1734 LeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlaThr 1753
Db 5251 CTGTCCCAAGTGCACACCTGCGCTGCTGGTCCCAACGAGCAGGACCCCTTGCAC 5310
Qy 1754 AlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSer 1773
Db 5311 GCCATCGACCGCTTGCCTACCTCCCACTGCGCCCGCCCACTTACGAGCGCCCAAGT 5370
Qy 1774 SerSerProLeuSerProGlyGlyProThrHisLeuThrIshisProThrThrThrSer 1793
Db 5371 AGCTCACCGCTGTCCCAGAGGAGCCCACTCACCTAGCTAAACCACTGCCATCTTCA 5430
Qy 1794 SerGluArgGluArgAspArgGluArgAspArgGluArgGluArgGluLysSer 1813
Db 5431 TCGAGCGGGAACGGGAACGTGAGCGGGAACGAGAC-----AAGTCC 5472

source	1. .8544	/organism="Mus musculus"	21	ProHisSerLeuSerTyProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu	40
		/mol_type="mRNA"	220	CCCCATGGCATCTCTACCGGTGAGTAGCCCGGTCCACACGAGCGTGGGGTGCCTT	279
gene	1. .8544	/db_xref="taxon:10090"	41	GlutylGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleleGln	60
CDS	160. .7548	/gene="Smrte"	280	GAGTACCAACACACCCCGTGTACTACCTCACACCTGTCCACCGGTTCATCATCCAG	339
		/gene="Smrte"	61	ProGlnArgArgProSerLeuSerGluPheGlnProGlyAsnGluArgSerGln	80
		/function="transcriptional corepressor"	340	CCACAGAGAGCGCCCTCACTGCTGTGAGAGTTCACGCTGGAGTGAACGGTCTCAG	399
		/notes="SMRte"	81	GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet	100
		/codon_start=1	400	GAGCTCCACCTGGCCCTGAGTCCCGCAGCTTCTGCTGAGCTGGGCAAGCCGACATA	459
		/product="silencing mediator of retinoic acid and thyroid hormone receptor extended isoform"	101	GluPheIleGluSerIleArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro	120
		/protein_id="AAB22972.1"	460	GAATTCACCGAGCAAGCGCCCGCTGGAGCTACTACCCGATACCTGCTGCGCCCA	519
		/db_xref="GI:4559296"	121	SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer	140
		/translations="MGSTQPVQAOTWRAEPYPHGISYIPVQIARSHDTDVLLEVOH	520	TCACCCCTCTGCGCCTGCGGAGCGAGTGGTCTGAAGACCTTACCAAGGACCGTAGC	579
		HPDYTHSLUSPGLTIQRRRPSLLSEFPQGSERSQELHLPESRTFLPELKGDEL	141	LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu	160
		TSKRPLLEPDLTLPRLPGLATQSGSDELTKDRSLAGLKEPSPSPSPHADPEL	580	CTGGCAGGCAAGCTGGAGCTGTGTGTACCTCCAGTCCCGCCGACGCTGACCTGAGCTA	639
		ELAPSLKEELITQNRDLVREITWVEQIISKLKKQQQLEEAKEAPEKPVSP	161	GluLeuValProProArgLeuSerIleGluLeuLeuIleGlnAsn---MetAspArgVal	179
		PIESKRSIVQIITDENRKAEEAARHLEGLGPOVELYNOPTDTRQYHENIKINQA	640	GAGCTGGCCCATCTCGACTGTCCAGAGGAGAGTGTATCCAGAACAGATTTGACCCCGTG	699
		MRKLLYFKRRNARKEQORFCQYDOLMEAWKKVERIENPRRAKESKVRYY	180	AspArgGluIleThrMetValGluGlnGlnIleSerLysLysLysLysGlnGlnGln	199
		EKQPFIRQRELQRMQRVQSGGLSMAARSSEHYSEIIDGLSELENLEKMRQ	700	GACCGTGAGATCACATGTGTAGCAGCAGATCTCCAGCTGAGAGAGAGCAGCAACAG	759
		LAVIRHVRDQRIKFINNGLMDPMKYIQVQVNNWMSQERDTFREKFMQHPKQ	200	LeuGluGluGluAlaLysProGluProGluLysProValSerProProProle	219
		FGIASFLEKTVAECLVYLYTKQNEYSKLSYRRSGKSGQQQQQQQQQQQQ	760	TTGGAGAGGAGGCGCCCAAGCGCCGCAACCGAGAGCTGTGTGCGCCACCCACCAT	819
		ARSQBEKEKEKEKEKEKEKEKEKEKEKEKEKEDTDSGDEKEAVASKRK	220	GluSerLysHisArgSerLeuValGlnIleTyrAspGluAsnArgLysLysAlaGlu	239
		TANSQRRKGRITRSMANANHEETATPOQSELSAMENNESRMTTEEMETAKGILL	820	GAATCAACACCGAGCGCTGTGCTCCAGATCATCTACGATGAGAACCGGAGAAAGCCGAA	879
		EHRNWSALTARWYKSTVQCKNFYNYKKRONLDEILOQHLKMEKERNARRKKKT	240	AlaAlaHisArgIleLeuGluGlyProGlnValGluLeuProLeuTyrAsnGln	259
		PAASSETAPFAAEDEMEASGANEELAEAEASQASNEVPRVCEGCPAAVN	880	GCCGCAACCGGATCTTAGAGGCTTGGGCGCCCGAGTGGAGCTGCTCTGTACCAACAG	939
		NSSDTSVSPSEATKQDKPTGTEALPAATQPPVPPPEPAAAPADSPVPDASG	260	ProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLys	279
		PPSPSPSPAPATVDQDEAPAPAPQTEDEAKEQSEAEIDVKGPEPSEAE	940	CGGTCTGACACACCGCAGTACCATGAAACATCAAAATAAACCCAGCGCATGCGGAAGAAG	999
		PPSPSKDHEEPEEKAKTEAETVSEAPLKVBEAGKAAVTKGSSSGATQDS	280	LeuIleLeuTyrPheLysArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCys	299
		DSSATCDEVDPEGGDKRLSPRLITPAGDPRASTPOKPLDLKQLKORAAAI	1000	CTGATCTTGTACTTTTAAGCGGAGAACCCAGCGCGCAAGCAGTGGGAAACAGCCCTCTGC	1059
		PELVTVHPPREDTVPKVPVPPPTOHLQPEGDVSOQSGSPRGKSRSPVPPAEK	300	GlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsn	319
		EAKPAFFAPFEGQSYRLSPHAGHRLSHPPREVITKSTRADPLFSTVPGHPLPL	1060	CAGCGTATGACCGAGCTCATGGAGGCTGGGAGAGAGGTAGAGCGCATAGAACAT	1119
		GLHDSARPLPPIINPPLISSAKHGLVLERQLGAIQSGMSVQLRVHSEHAKPMG	320	ProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGlu	339
		PLTMELFLAVDPKLLTALAPPVEASPRASQYPGCRPQLQLXHPHPTADVLYKGT	1120	CCGCAAGAGGAGGCGCAAGGAGCAAGGTGAGGAGTACTTACGAGAAACAGTTTCCCGGAG	1179
		ISRIVGDSPLRLDAREDTLPKHVIYEGKGHVLVSYEGGMSVSCSKEDSRSSGP	340	IleArgLysGlnArgGluLeuGlnArgMetGlnSerArgValGlnGlnArgGlySer	359
		PHETAAPKRTYDMWGRVGTVTASIEGLMGRAIPEQHSPLKEQHIRGISTQIP	1180	ATCCGCAAGCGCGGAGCTGCGAGGAGCGCATCAGACGAGGTTGGCCGCGCTGGCAGT	1239
		RSVVEAQDYLRKALKKEGTPPPPPPRDLITETYPKRPDLPLGLPLKPTHEGVV	360	GlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleleAspGlyLeu	379
		ATVKEAGRSIHETPRELRTPELPLAPRLKGSITQCTPLKYDSGAPSTGTXKHIV	1240	GGGCTCTCCATGTCTCGGCTGCCCGACGTGAGCATGAGGTTTCTGAGATCATTGATGCTTG	1299
		RSIIGSPFPFPAHDLTMADALALERACTEESLSRSGTSSGAGSITRGAAPVVP			
		ELGKQSPPLTYDHGAPFTSHLPSPVPTRETPRLQEGSLSSKASODRKLITSTP			
		REIAPKSPHTYHHPHTSPYHLIRGVTGDLVYRCHPLAFDPTSIPRGTPLEAAA			
		AAVYLPHLAPSTPHLYPYLIRGVPDTAALENQTIINDYISQQMHNAASAMA			
		QADMLRGLSPRESSIALNYSAGPGLIDLSQVPLPLVLPVPTGPATADRLAFLP			
		TAPPPSSHSSSPLSPGCTHLAKPTATSSSERERERERERDKSLTSTTVVEHAP			
		WRFGTQSSGAGSSRSPASHOHSFISPKTQALQORSPVLNMTSMKGVTISVEGT			
		PTVLRSTSSSPVPAATFPFATHCPGLGTLEGVYPTLMEPVLPLPKETSRKVARPERP			
		VDGHAFLTKPPAREPASSPKSEPRSLAPPSSHTAARTPAKSLAPHASPDPPG			
		PTGASDLHREKTKQKPSIOLELRSLGVHSGAGYSPDVEPTSPVSSPSLTHDKGLS			
		KPLELEKSHLEGLRHKQPMKLSAEAAHLPLRLPESQSPSPLQTPAICIKGH			
		QRVVTLAQHISEVITQDYTHHPQQLSGPLPAPLYSPGASCPVLDLRPPSLYLGP			
		PDGTFARGSPHSGGKSPKSPKSVLGSSEDAIEPVPPEGMPTEPHGARSAYPLL			
		YRDGQEQEPMGSKSPGNTQPPAFKSLTESAMVKSKKKINKKLNTHNRNPEY			
		NIQCGTEIFNPAITAGMLTCSRQAVQSHASTNMGLEAIRKALMGKYDQWEEPPP			
		LGNANFNPLNASLAPAAAMPTIATDGRSHALTSFGGKAKVSRGSRSSKAPAP			
		GLASGDRPPSVSVSHSGDCNRRPLTLNVRVEDRPSAGSTPPYPLNMLRLQAVMA			
		SPPPGLAAGSGPLGPHHAWDEBPRLCCSYETILSDSE"			

ORIGIN

Alignment Scores:					
Pred. No.:	3,11e-211	Length:	8544		
Score:	10832.50	Matches:	2149		
Percent Similarity:	87.94%	Conservative:	90		
Best Local Similarity:	84.41%	Mismatches:	197		
Query Match:	81.97%	Indels:	117		
DB:	10	Gaps:	33		

US-09-522-753-5 (1-2517) x AF125671 (1-8544)

Qy	1	MetSerGlySerThrGlnLeuValAlaGlnThrTyrArgAlaThrGluProArgTyrPro	20
Db	160	ATGTGAGGATCCACACGCTGTGGCAGACATGCGCGGCTGCTGAGACCCCGCTACCCA	219

Db 6292 TACCCCTCATGGAGCCGCTGTTTACCAAGGAGACCTCTCGGGTCGCCGCCGAGCG 6351
QY 1950 gProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuG1 1970
Db 6352 GGCCCGGGTGGAGCGTGGCATGCCCTTTTACCAAAACCCCGGGCCGG-----GA 6402
QY 1970 uProAlaSerSerProSerLysGlySerGluProArgProLeuValProValSerG1 1990
Db 6403 GGCCGCTCTCTACCCAGCAGAGCTCCGAGCCCGCATCCCTAGCACCCCGGCTCCAG 6462
QY 1990 yHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAs 2010
Db 6463 CCACACAGCATGCGCCGACCCAGCAAGAACCTTGACCCCAACCATGCCAGTCCGGA 6522
QY 2010 pProProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPr 2030
Db 6523 CCGCCCGGGG---CCACCTCGGCCCTCAGATGTGCACCGAGAAAGACATCAAGTANACC 6579
QY 2030 oPheSerIleGlnGluLeuLysArgSerLeuGlyTyrHis---GlySerSerTyrSe 2049
Db 6580 CTTTTTCATCCAGGAATTGGAATCCGTTCTCTGGGTATACACAGTGGAGCTGGCTACAG 6639
QY 2049 rProGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysG1 2069
Db 6640 CCGCATGGGTGGAGCCCATCAGCCCGGTGAGCTCCCGCAGCTCACCACGACGACAGGG 6699
QY 2069 yLeuProLysHisLeuGluLeuLysAspLysSerHisLeuGluGlyLysLeuArgProLy 2089
Db 6700 GCTCTCCAAACCTCTGGAGAGCTAGAGAGAGCCCACTTGGAGGGGAGCTGGCGACAA 6759
QY 2089 sGlnProGlyProValLysLeuGlyGlyGluAlaHisLeuProHisLeuArgProLe 2109
Db 6760 GCAGCCAGGCCCATGAGCTCAGCGCGGAGGCTGCCCATCTCCACATCTGCGGCCACT 6819
QY 2109 uProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHi 2129
Db 6820 GCCCGAGAGCAGCCCTCATCCAGCCCACTCTCTCCAGACTGCGCCCAAGGATCAAGGTCA 6879
QY 2129 sGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrAr 2149
Db 6880 CCAGAGGGTGTCACTGCTGCTCAGCACATCAGCAGAGTCAATTACCAGAGACTTACACGG 6939
QY 2149 gHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAl 2169
Db 6940 CCACACCCGCGCAGCTCAGTGGCCCTTCCCGCCCTCTCTACTCTCTCTCTCTCTCTCT 6999
QY 2169 aSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAs 2189
Db 7000 CAGCTGCCCTGTCTCTGGATCTTCCGCGCCCAACCCAGTGCCTTACTCTCCACCCCGCA 7059
QY 2189 pHisGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluPr 2209
Db 7060 CCATGGACCCCGAGCGGGATCCCGCCCACTGAGGGGGCAAGAGTCCCAAGACC 7119
QY 2209 oAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProG1 2229
Db 7120 CAGCAAAACATCGTCTCTGGCAGCAGCAGGATGCATTTAGSCCTGTGTCTCCACACAGA 7179
QY 2229 uGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspG1 2249
Db 7180 GGGCATGACTGAGCCAGCATGCTCGGAGCACTGGGTATCCCACTGCTGTATCGAGACGG 7239
QY 2249 yGlnGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProPr 2269
Db 7240 GGAACAGGGCGAGCCCC---AGGATGGTCTAGAGTCTCCAGGCAACACAGCCAGCCGCC 7296
QY 2269 oAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnG1 2289
Db 7297 AACCTTCTTCAGTAAGCTGACTGAGAGCACTCGGCCATGGTGAAGTGAAGAGCAGGA 7356
QY 2289 uIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnPr 2309

Db 7357 GATCAACAAGAACTCAACACCAACCGGAACCGAGCCAGAAATACAATATTGGCCAGCC 7416
QY 2309 oGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSe 2329
Db 7417 TGGGACGGAAATCTTCAACATGCCGCCATCACTGGAGCAGGCCCTATGACCTGTAGAAG 7476
QY 2329 rGlnAlaValGlnGlnHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAl 2349
Db 7477 CCAGGCGGTGCAAGAACACGCGCAGCAACCAATGGGCTAGAGCCCATTTATTAGAAGGC 7536
QY 2349 aLeuMetGlyLysTyrAspGlnTrpGluLysSerProProLeuSerAlaAsnAlaPheAs 2369
Db 7537 ACTATGGGTAATATGATCAGTGGGAAGAGCCCGCCGCTCGGCGCCAAATGCTTTTAA 7596
QY 2369 nProLeuAsnAlaSerAlaSerLeuPro---AlaAlaMetProIleThrAlaAlaAspG1 2388
Db 7597 CCTCTGAATGCCAGCGCAGTCTGCCGCTGCTGCTATGCCCATAAACCACTGCTGACGG 7656
QY 2388 yArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyAr 2408
Db 7657 ACGGAGTGACCAACGCACTCCTCGCCAGGTGGAGTGGGAAAGCCCAAGGTCTCTGCGCAG 7716
QY 2408 qProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProPr 2428
Db 7717 ACTAGACGCCGAAAGCCCAAGTCCGCGACACAGGCTTAGGCTCCGGAGACCGACCCCC 7776
QY 2428 oSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnAr 2448
Db 7777 TTCTGTCTCTCAGTACACTCAGAGGGGACTGCAATCGCCGACACCACTCACCAACCG 7836
QY 2448 gValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIl 2468
Db 7837 TGTGTGGGAGGACCGGCCCTCATCTCGAGGGTCCAGGCCATTCCTCCCTACACCTTTGAT 7896
QY 2468 eMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySe 2488
Db 7897 TATGAGGCTACAGGAGGAGTGTATGGCTTCCCGCCCGCCCTGCGGAGGAGCAG 7956
QY 2488 rGlyProLeuAlaGlyProHisHisAlaTrpAspGluGluProLysProLeuLeuCysSe 2508
Db 7957 CCGGCCCTTAGTGTGTCTCCACCGCTGGGATGAGGAGCCCAAGCCCACTGCTGTGTTTC 8016
QY 2508 rGlnTyrGluThrLeuSerAspSerGlu 2517
Db 8017 ACAGTATGAGACACTCTCGACAGCGAG 8044
RESULT 4
AF125671
LOCUS
DEFINITION
Mus musculus silencing mediator of retinoic acid and thyroid hormone receptor extended isoform (Smrte) mRNA, complete cds.
ACCESSION
AF125671
VERSION
AF125671.1
KEYWORDS
GI:4559295
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
REFERENCE
1 (bases 1 to 8544)
Park,E.J., Schroen,D.J., Yang,M., Li,H., Li,L. and Chen,J.D.
SMRte, a silencing mediator for retinoid and thyroid hormone receptors-extended isoform that is more related to the nuclear receptor corepressor
Proc. Natl. Acad. Sci. U.S.A. 96 (7), 3519-3524 (1999)
JOURNAL
MEDLINE
99199215
PUBMED
10097068
REFERENCE
2 (bases 1 to 8544)
Chen,J.D.
Direct Submission
Submitted (03-FEB-1999) Pharmacology and Molecular Toxicology,
University of Massachusetts Medical School, 55 Lake Avenue North,
Worcester, MA 01655, USA
JOURNAL
TITLE
Location/Qualifiers
FEATURES

[illegible]

3166	DB	GCACAGAGGCCCGAGGCGCTCTGAGGAGCGCCCGGAGAGGTGTAAAGAGTGACCAAGGGA	3222
868	QY	uGluAlaGluGluGlyProAla---LysGlyLysAspAlaGluAlaAlaGluAlaThrAl	887
3226	DB	GGAGCCGAGGAGAGCGCTGAACACAAAGCCAGGGCACAGAGCCATTGAACTGTGTGTC	3285
887	QY	aGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThr---	906
3286	DB	TGAGGCACCACTTAAGGTGGAG-----GAGGCTGTGTAGCAAGCGACTGTGACCAAGGG	3339
906	QY	aLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGluVa	926
3340	DB	TTCAGCTTCAGGTGCCACCCAGACAGCTGACTTCAGTGCCACCTGCAGTGC CGATGAGGT	3399
926	QY	lAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeuLeuTh	946
3400	DB	GGACGAACCGAAGGAGTGCACAAAGCGAGCTGTCTACCAAGGCCAGCCTCTCTCAC	3459
946	QY	rProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLe	966
3460	DB	CCCGGCTGGAGATCCCGGCGCACTACTCGCCGCCAGAACCGCTGGACCTGAAGCAGCT	3519
966	QY	uLysGlnArgAlaAlaAlaLileProProLileGlnValThrLysValHisGluProProAr	986
3520	DB	GAAGCAGCAGCAGCGCCCATCCCTCCCTATCCAGGTCACCAAGTCCATGAGCCCCCCC	3579
986	QY	gGluAspAlaAlaProThrLysProAlaProAlaProProProGlnAsnLeuGl	1006
3580	DB	GGAGGACACAGTACCCCAAGCAGTTCCTCCCTGTGCTCCACCCAGCAGCACCTACA	3639
1006	QY	nProGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerPr	1026
3640	DB	GCCAGAGGCTGACGTGTCTCAGCAGTCGGGAGGAAGTCCACGTGGCAAGTCCCGCAGCCC	3699
1026	QY	oAlaProProAlaAspLysGlu-----AlaPheAlaAlaGl	1038
3700	DB	AGTGCTCTCTGCCGAGAAAGGAGGAGAACCCGCAATCTTTCCGGCTTTTCCCACTGA	3759
1038	QY	uAlaGlnLysLeuProGlyAspProProCysTTPThrSerGlyLeuProPheProValPr	1058
3760	DB	GGGCCCCAAGCTACCGACTCAGCCCCCACCGCTGGTCATCGGGCTGCCTTCCCATCCC	3819
1058	QY	oProArgGluValLileLysAlaSerProHisAlaProAspProSerAlaPheSerTyAl	1078
3820	DB	TCCAGGGAGGTGATCAAGACTTCCCCACACGCGCTGACCCCTCTGCCTTCTCCTACAC	3879
1078	QY	aProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProAr	1098
3880	DB	ACCCCCGGTCACCGCTGCTCTGGGCTTCCACGATAGTGC CGGCGCCCTGCTGCCACG	3939
1098	QY	gProProThrIleSerAsnProProProLeuLileSerSerAlaLysHisProSerValle	1118
3940	DB	TCCCCC---ATCTTAACCCCCACCCCTCATCTCTCTGTGCCAGCATCTCCGGCGTACT	3996
1118	QY	uGluArgGlnLileGlyAlaLileSer---GlnGlyMetSerValGlnLeuHisValProTy	1137
3997	DB	TGAGAGCAGCTGGGTGCCATCTCCACGACAGGGGATGTCAGTCCAGCTTCGTGTGCGCTCA	4056
1137	QY	rSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAs	1157
4057	DB	CTCAGACATGCCAAGGCCCCCATGGGCCCTCTCACCATTGGGGCTGCCCTTGCCGTGA	4116
1157	QY	pProLysLysLeuAlaProPheSerGlyValLysGlnGluLeuSerProArgGlyGl	1177
4117	DB	CCCTAAGAACTG-----	4129
1177	QY	nAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGl	1197
4130	DB	-----GG	4131
1197	QY	YThrAlaLeuGlySerValProGlyGlySerLileThrLysGlyLileProSerThrArgVa	1217
4132	DB	GACAGCACTGGGCTCGCCACCAAGTGGAAAGCAATCACCAAGGGCTCCCCAGTACCGGGC	4191

```
NSDTSVSPSPSEAMKOTGPKPTGTGTGTEALPAATQPPVPPPPPPAVAPAEPSVPDASG
PPSPESHLPRLWLTMNKKRLLQLPRMPRSRLRRRSMWKEPPEASEE
PPSVKSDHEEPEEDKXGTEAETVSEAPLKVEAGSKAAVTGSSSGATQDS
DFSATCSADVEDEGGDKRLSPRLSTPAGDPRASTSPKQLDLKOLKQRAAI
PPIQTVKHEPPREDIVPPKVPVPPPTLOHQEDGVSOQSGSPRGKSRSPVPAE
KEAEKAPFAPEPTGPKLPPTPPWSSGLPFIIPREVIKTSPhAADPSAFSPYPPG
HPLPLGHDASVPVLPPIPNPPPLISSAKHPGVLERQLGALSQOQMSVOLRVPHSE
HAKAPMPLTMGLAVDPKLGTLGTSATSGSIITKGLPSTRAADGSPSTRGSIHTGTP
ADVLVYGTTSIRIVGEDSPRLDRAREDITLPKGVITYEGKKGHVLSYEGMSVSQCSKE
DGRSSGPPHETAPAKRTYDMREKRLKREGTPPPPPDRDLTETVKRPLDLPLGLKL
GSITQGISYVEAODYLREAKRLKREGTPPPPPDRDLTETVKRPLDLPLGLKL
GPLEAAAAAYILPHLASPTYPHLYPPYLIRGPDPTAALENROTIINDYITSQOH
HNAASMAORADMLRGLSPRESSLALNYAAGPRGIDLSQVPHLPVLVPPPTGTPATA
IDRLAYLPTAPPFPSSSRHSSPLSPGSPHAKPTATSSSERERERERERDKSILTST
TTVEHAPTRPQSGSGAGSSRPASHTHQSPISPRTPQDALQOQPSVLNHTSMKGV
TVSEPTPTVLRWAKSTSTSPVPPAAITPPATHCPGLGTLEGVYPTLMFVPLPKE
TSVRPERARVADGHAFLTKPPREPAPSPSKSEPRSLAPPSSHTALARTPAKNL
APHASDPDPAPTASDLHREKTKQSFSTQOELERSLGSHGAGSYSDGVDFPISVS
SPSLTHDKGLSKLELEKSHLQKOPGPMKLSAEEAHLPHLRPLPESQPSSP
LLOTPAGIKGHQVVTLAQHSI SEVITQDYTRHHPOOLSGPLPAPLYSFPGASCPVLDL
RRPSPDYLPPPHGTARGSPHSEGGKSPSPKTSVLGSSSEDAIEPVSPPEGMTEP
GHARSTAYLLYRDGEGQEPRMGLSEFGNTSQPTTFKSLTESNANWKSKEOENKK
LNTNRNEPEYNTIQGPTEIFPNASAITAGMLTCRQAOVHSHASTNMGLEAIRKALM
KTYDWEPEPTPLAGNAPFNPLNASILPAAAMPTITADGSDSHALTSPGGGSGASVGR
PSRKAKSPAPGLASGRPPSPSVSHSEGDGCRNRTPLTNRWEDRESSAGSTPPFYNP
LINELQAGVNASPPPPGLAAGSGPLAGPHHAWDEEPKPLLCQSYETLSDBE"
731...1387
misc_feature
/ note="Region: deleted in SMRT beta isoform encoded by
GenBank Accession Number AF113002"
```

ORIGIN

```
Alignment Scores:
Pred. No.: 5,54e-216 Length: 8388
Score: 11067.00 Matches: 2160
Percent Similarity: 88.20% Conservative: 89
Best Local Similarity: 84.71% Mismatches: 192
Query Match: 83.75% Indels: 110
DB: 10 Gaps: 29

US-09-522-753-5 (1-2517) x AF113001 (1-8388)
QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20
Db 626 ATGTCAGGATCCACAGCGCTGTGCGCACAGACATGCGCGGCTGTGAGCCCCCGCTACCCA 685
QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
Db 686 CCCCATGGCATCTCTACCCCGGTGAGATAGCCCGTCCCAACAGCGAGCTGGGCTGCTT 745
QY 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
Db 746 GAGTACCAACACACCCCGTGTACCTACACCTCACACCTGTCAACCGGTTCATCATCCAG 805
QY 61 ProGlnArgArgProSerLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
Db 806 CCACAGAGAGAGCGGCGCTCACTGTGTGAGAGTTCCAGCTGGGAGTGAACGGTCTCAG 865
QY 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuLeuGlyLysSerGluMet 100
Db 866 GAGCTCCACCTCGGCCCTGAGTCCCGCAGCTTCCTGCTGAGCTGGGCAAGCCCGACATA 925
QY 101 GluPheIleGluSerIysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
Db 926 GAATTCACCGAGAGCAAGCGCCCGCTGAGCTACTACCCGATACCTGCTGCGGCCA 985
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
Db 986 TCACCCCTCTGCGCCACTGGGCGAGCGCGAGTGGGTCTGAAGACCTTACCAAGGACCGTAGC 1045
```

```
141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
1046 CTGCAGGCAAGCTGGAGCGCTGTGTACCTCCAGTCCCCCGCACGCTGACCCCTGAGCTA 1105
161 GluLeuValProProArgLeuSerLysGluLeuLeuIleGlnAsnMetAspActValAsp 180
1106 GAGCTGGCGCATCTCCGACTGTCCAAGGAGGAGTGATCCAGAACATGGACCGCGTGAC 1165
181 ArgGluIleThrMetValGluGlnGlnIleSerLysLysLeuLysLysGlnGlnGlnLeu 200
1166 CGTGAGATCACCATGTGTAGACGACAGATCTTCCAAGCTGAAGAAGACGACCAACAGTTG 1225
201 GluGluGluAlaAlaLysProGluProGluLysProValSerProProProIleGlu 220
1226 GAGGAGGAGGCGCGCAAGCCCGCAACCCGAGAAGCGTGTGTGCGCCACCAACCATAGAA 1285
221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysLysAlaGluAla 240
1286 TCAAAAGCACCGAAGCCTGTGTCAGATCATCTACGATGAGAACCGGAGAAAGACCGAAGCC 1345
241 AlaHisArgGluLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
1346 GCACACCGGATCCTAGAAGCGCTGGGGCCCCAGGTGGAGCTGCCTCTGTACAAACGACCG 1405
261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
1406 TCTGACACACGCGCATCATCAATGAAACATCAAAATAAACACGAGCGCATCGGAAGAAGTG 1465
281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGln 300
1466 ATCTTGTACTTTAAGCGGAGGAACCAACGCGGCAAGCAGTGGGAAACAGCGCTTGTGCCAG 1525
301 ArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnPro 320
1526 CGCTATGACCACTCATCGAGGCGGTGGGAGAGAAGGTAGAGCGCATAGAGAACCAATCGG 1585
321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
1586 CGAAGGAGGCGCAAGGAGAGCAAGGTGAGGGAGTACTACGAGAAACAGTTCCCGGAGATC 1645
341 ArgLysGlnArgGluLeuGlnGlnArgMetGlnSerArgValGlyGlnArgGlySerGly 360
1646 CGCAAGCAGCGGAGCTGCGAGGCGCATGCGAGCAGCGGTGGCGGCGGCGTGGCAGTGGG 1705
361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
1706 CTCTCCATGTGCGCTCGCCGCGAGTGAAGCATGAGGTTTCTGAGATCATTTGATGGCTTGTCT 1765
381 GluGlnGluAsnLeuLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
1766 GAGCAGGAGAACCTGGAGAAGCAGATGCGCCGAGCTGCGCGCTGATCCCGCCCATGTTGTAC 1825
401 AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
1826 GAGCGCGACACGACGAGGATCAAGTTTCATCAATGGAATGGAATGATGATGATGATGATGAT 1885
421 LysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGlnLysGluThrPhe 440
1886 AAGGTCTTACAGGACCGTCAAGTTTACCAACATGTGGAGGCGAGGAGGAGGACACCTTC 1945
441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
1946 CGTGAGAGAGTTTATGACGACACCTTAAGAACTTTGGCTGATTCCTTCCTTCCTGGAGAGA 2005
461 LysThrValAlaAlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
2006 AAGACGCTCGCTGAGTGTGCTCTTATTACTACTACCAAGAAAGAAATGAAATTAACAAG 2065
481 SerLeuValArgArgSerTyrArgArgGlyLysSerGlnGlnGlnGlnGlnGln 500
2066 AGCTTGTGTGAGGCGGAGCTATCGCGCGCGTGGCAAGGCGAGCAGCAGCAGCAGCAGCAA 2125
501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluLys 520
```

Db 6282 GTGGAGCCCGTGTGAGCTCACCAGTCTGACCCACGACCAAGGGGCTCCCCAA 6341
Qy 2072 sHisLeuGluLeuAspLysSerHisLeuGluGlyLeuLeuArgProLysGlnProGl 2092
Db 6342 GCACCTGGAAGAGCTGACACAGAGCCACCTGGAGGGGAGCTGCGGCCACAGCAGCAGG 6401
Qy 2092 yProValLysLeuGlyGlyGluAlaHisLeuProHisLeuArgProLysGlnProGluSe 2112
Db 6402 CCCCGTGAAGCTTGGCGGGGAGCGCCGACCTCCACACCTCGCGCGCTGCTGAGAG 6461
Qy 2112 rGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVa 2132
Db 6462 CCAGCCCTCGTCCAGCGCGCTGCTCCAGACCGCCCGAGGGGTCAAGGTCACACAGCGGGT 6521
Qy 2132 lValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisIlePr 2152
Db 6522 GGTCAACCTGGCCAGCACATCATGTGAGTTCATCACAGGACTACACCGGACCAACCC 6581
Qy 2152 oGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysPr 2172
Db 6582 ACAGCAGCTCAGCGCACCCCTGCGCGCCGCTCTACTCTCTCTGCGGCGAGCTGCC 6641
Qy 2172 oValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAl 2192
Db 6642 CGTCTGGACCTCCGCGCCACCCAGTGACCTTACCTCCCGCCCGGACCATGGTGC 6701
Qy 2192 aProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysTh 2212
Db 6702 CCCGCGCGTGGTCTCCCGCCAGCAGCAGGCGGAGAGTCTCCAGAGCCAAACAGAC 6761
Qy 2212 rSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetTh 2232
Db 6762 GTCGCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 6821
Qy 2232 rGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnTh 2252
Db 6822 GGAGCCAGGCGACTCCCGGAGTCTGTGTATCCCGCTGTGTATCCCGGATGGGAACAGAC 6881
Qy 2252 rGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePh 2272
Db 6882 GGAGCCAGCAGAGATGGCTCCAGTCTCCAGCGACACACAGCAGCGCGCGCGCTCTCT 6941
Qy 2272 eSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGlnLysLeu 2292
Db 6942 CAGCAAGCTCAGCAGAGCAACTCCGCGCATGGTCAAGTCCAGAGCAAGAGAGATCAACA 7001
Qy 2292 sLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGl 2312
Db 7002 GAAGCTGAACACCCACACCGGAATGAGCCTGAATATATATATATATATATATATATAT 7061
Qy 2312 uIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVa 2332
Db 7062 GATCTTCAATATGCGCCCATCACCGGACAGCGCTTATGACCTATAGAGCCAGCGCGGT 7121
Qy 2332 lGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGl 2352
Db 7122 GCAGGAACATGCCAGCACCAATGCGGCTGGAGGCCATAATATAGAAAGGCACTCATGG 7181
Qy 2352 yLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAs 2372
Db 7182 TAAATATGACACAGTGGAGAGTCCCGCGCGCTCAGCGCCAAATGCTTTTAAACCTCTGA 7241
Qy 2372 nAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHi 2392
Db 7242 TGCAGTGCCAGCGCTGCCCTGCTATGCCATTAACCGCTGCTGACGAGCGAGTGAACCA 7301
Qy 2392 sThrLeuThrSerProGlyGlyGlyLysAlaLysValSerLysGlyArgProSerSerAr 2412
Db 7302 CACACTCACCTCGCCAGTGGCGGCGGAGGCGCAAGTCTCTGGCAGACCCAGCAGCGG 7361
Qy 2412 gLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSe 2432
Db 7362 AAAAGCCAGTCCCGCGCGCTGGCATCTGGGAGCCCGGCGCCAGCTCTCTCTCTCTCT 7421

Qy 2432 rValHisSerGluGlyAspCysAsnArgThrProLeuThrAsnArgValTrpGluAs 2452
Db 7422 AGTGCACTCGGAGGAGAGCTGCAACCGCGGAGCGCTCACCAACCGCGTGTGGGAGA 7481
Qy 2452 pArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGl 2472
Db 7482 CAGGCGCTCGTCCGAGGTTCCACGCCATTCCTCCATACAAACCCGCTGATCATGGCGTGA 7541
Qy 2472 nAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeuAl 2492
Db 7542 GCGGCTGTGATGGCTTCCACCCCGCGGCTCCCGCGGAGCGGCGCCCTCGC 7601
Qy 2492 aGlyProHisHisAlaTrpAspGluGluProLysProLeuLeuCysSerGlnTyrGluTh 2512
Db 7602 TGGCCCCCACCACGCTGGGAGGAGGAGCCCAAGCCACTGCTCTGCTCGCAGTACGAGAC 7661
Qy 2512 rLeuSerAspSerGlu 2517
Db 7662 ACTCTCCGACGAGGAG 7677
RESULT 3
AF113001
LOCUS
DEFINITION
Mus musculus silencing mediator of retinoic acid and thyroid hormone receptor alpha mRNA, complete cds.
ACCESSION
AF113001
VERSION
AF113001.1
KEYWORDS
GI:4454547
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 8388)
Ordentlich,P., Downes,M., Xie,W., Genin,A., Spinner,N.B. and Evans,R.M.
Unique forms of human and mouse nuclear receptor corepressor SMRT
Proc. Natl. Acad. Sci. U.S.A. 96 (6), 2639-2644 (1999)
JOURNAL
MEDLINE
99178941
PUBMED
10077563
REFERENCE
2 (bases 1 to 8388)
Downes,M.R., Ordentlich,P. and Evans,R.M.
Direct Submission
Submitted (11-DEC-1998) Gene Expression Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA
FEATURES
source
Location/Qualifiers
1..8388
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/tissue_type="spleen; brain"
626..8047
/function="transcriptional co-repressor"
/note="MSMT alpha"
/codon_start=1
/product="silencing mediator of retinoic acid and thyroid hormone receptor alpha"
/protein_id="AAD20944.1"
/db_xref="GI:4454548"
/translation="MSGSTQPVATQWRAEAPRYPHGISYVPQVIARSHDTDVGLLVQHH
HPRDVTSLSPGSIIPQRRPRLSSEFQSGSERSOELHRLPESRTFLPELKGPDIEF
TSKPRLELLPTLLRSPPLATGQPSGSDLTDRSLAGLKPVSPPSPHADPEL
ELASLKEELIQNDVDRITWVEQIISLKKQQLEREAAKPPPEKPVSP
ISKRSLSVQIYIDENRKAEEAHLLEGQVPELVLYNQSDTFQHENIKNOAM
RKXLLYFKRRNHAKQWRFQCDLMEAWKVERIENNRRAKESKREYIE
KQFPIRKQRELMQSRQSGSLSSAARSEHVESEIIDGSEQENLQKMRQL
AVIPMLYADQQRIFKFNMGIMDDPMKYQDRQVTNWSQEDRTFFEKPMHPKN
FGLIASLKEETVAECVLYLTQADENYKSLVRSYRRRGSQOQOQOQOQOQ
ARSSQEEKEEKEADEKQDAENEKEELSKETDTSGENHEKEAVSKGRK
TANQORRRGRITRSMANHEETATPOOSSELASMENNESSRTEEMETAKGGLL
EHRNWSAIARVMGSKTSQCKNFYNYKRNQLDEILQOHLKMEKERNARRKKKT
PAAASEETAFPPAAEDEMEASGASANEELAEAEASQASGNEVPRVCGEPAVN

QY	1332	YArgAlaIleProProGluArgHisSerProHisHisLeuLeuGluGlnHisIleAr	1352
DB	4122	CCGTGCCATCCCGCGGAGCGACACAGCCCCCACCACCTCAAGAGCAGCACACCATCCG	4181
QY	1352	gGlySerIleThrGlnGlyIleProArgSerTyrrValGluAlaGlnGluAspTyrrIleuAr	1372
DB	4182	CGGGTCCATCACACAGGGATCCCTCGGTCTACTAGTGGAGGCACAGGAGACTACCTGCG	4241
QY	1372	gArgGluAlaIleLeuLeuLeuArgGluGlyThrProProProProProSerProSerArgAs	1392
DB	4242	TCGGGAGGCCAAGCTCTTAAGAGGGAGGCGACGCTCCCGCCCCCACCAGCCCTCACGGGA	4301
QY	1392	pLeuThrGluAlaTyrrLysThrGlnAlaLeuGlyProLeuLeuLeuLysProAlaHisG1	1412
DB	4302	CCTGACCGAGGCTTACAGACGCGAGGCCCTGGGCCCTTGAAGCTGAAGCGGCCCCATGA	4361
QY	1412	uGlyLeuValAlaThrValLysGluAlaG1ArgSerIleHisGluIleProArgGluG1	1432
DB	4362	GGGCTGTGGTCCACCGTGAAGAGGCGGCGCGCTCCATCATGAGATCCCGCGCGAGGA	4421
QY	1432	uLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLeuGluGlySerIleTh	1452
DB	4422	GCTCGCGCACACGCGCCGAGCTGGCCCTGGCCCGCGCGCTCAAGAGGGCTCCATCAC	4481
QY	1452	rGlnGlyThrProLeuLysTyrrAspThrGlyAlaSerThrThrGlySerLysLysHisAs	1472
DB	4482	GCAGGGCACCCCGCTCAAGTAACGACACCGGGCGGTCCACACCTGGCTCCAAAAGCAGCA	4541
QY	1472	pValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspVa	1492
DB	4542	CGTACGCTCCCTCATCGCACGCGCGCGGACGTTCCACCGCTGACCCCGCTGGATGT	4601
QY	1492	lMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrrGluGluSerLeuLysSerArgPr	1512
DB	4602	GATGGCGACGCGCGGCACTGGAAACGTGCTGTCTACGAGAGAGCCCTGAAGAGCGCGCC	4661
QY	1512	oGlyThrAlaSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProG1	1532
DB	4662	AGGACCGCCAGCAGCTCGGGGGCTCCATTGGCGCGCGCGCCCGCTATTGTCCTGA	4721
QY	1532	uLeuGlyLysProArgGlnSerProLeuThrTyrrGluAspHisGlyAlaProPheAlaG1	1552
DB	4722	GCTGGTGAAGCGCGGAGAGCCCTGACCTATGAGGACACCGGGCACCCCTTTGCGCG	4781
QY	1552	yHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluG1	1572
DB	4782	CCACCTCCACAGAGTTCCGCCGTGACCATGCGGGAGCCACCGCGCGCTGACGAGGG	4841
QY	1572	ySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluI1	1592
DB	4842	CAGCCTTTCTGTCACAGCAAGGCATCCAGGACCGAAGCTGACGTGACGCGCTGTGAGAT	4901
QY	1592	eAlaLysSerProHisSerThrValProGluHisHisProHisProIleSerProTyrrG1	1612
DB	4902	CGCCAAAGTCCCGGCACAGCACCGTGGCCGAGCACACCCCATCTCGCCCTATGA	4961
QY	1612	uHisLeuLeuArgGlyValSerGlyValAspLeuTyrrArgSerHisIleProLeuAlaPh	1632
DB	4962	GCACCTGCTTCGGGGCGTGAAGTGGCGTGAGCTGTATCGAGCCACATCCCTTCGGCTT	5021
QY	1632	eAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaIaTyrrTyrrLeuPr	1652
DB	5022	CGACCCCACTCCATACCCCGCGCATCCCTCTGGACGCGCGCTGCTACTACCTGCC	5081
QY	1652	oArgHisLeuAlaProAsnProThrTyrrProHisLeuTyrrProProTyrrLeuIleArgG1	1672
DB	5082	CCGACACGTGGCCGCCAACCCACCTACCCGACCTGTATCCACCTACTCATCTCCCGCG	5141
QY	1672	yTyrrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrrIleThrSe	1692
DB	5142	CTACCCCGACACGGCGCGCTGGAGACCGGACCATCATCATGACTACATCATCTC	5201
QY	1692	rGlnGlnMetHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgG1	1712

DB	5202	GCAGCAGATCCACCAACACGCGCCACCGCCATGGCCCGCAGAGCTGATATGCTGAGGG	5261
QY	1712	yLeuSerProArgGluSerSerLeuAlaLeuAsnTyrrAlaAlaGlyProArgGlyIleI1	1732
DB	5262	CCTCTCGCCCCGCGAGTCTCGCTGGCATCTAACTACGCTGCGGGTCCCGAGGATCAT	5321
QY	1732	eAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAl	1752
DB	5322	CGACCTGTCCCAAGTGCACACCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	5381
QY	1752	aThrAlaMetAspArgLeuAlaTyrrLeuProThrAlaProGlnProPheSerSerArgHi	1772
DB	5382	CACCGCATAGGACCGCTTGCCTTACCTCCCGCACCGCGCGCCCTTCAGCAGCGCGCA	5441
QY	1772	sSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSe	1792
DB	5442	CAGCAGCTCCCGACTCTCCCGCAGGAGTCCACACACTTGCACAAAACCAACACACGTC	5501
QY	1792	rSerSerGluArgGluArgAspArgGluArgAspArgAspArgGluArgGluGly	1812
DB	5502	CTCGTCCGAGCGGAGAGACCGCGATCGAGAGCGGACCGGGATCGGAGCGGGA	5561
QY	1812	sSerIleLeuThrSerThrThrThrValGluHisAlaProIleThrArgProGlyThrG1	1832
DB	5562	GTCCATCTCATCGTCCACACGACCGTGGAGCAGCACCCCATCTGGAGACCTGTGTACAG	5621
QY	1832	uGlnSerSerGlySerSerGlySerGlyGlyGlyGlyGlyGlySerSerSerArgProAl	1852
DB	5622	GCAGACGACGCGCAGCAGCGCGCAGCGCGGGGTGGGGGCGCAGCAGCGCCCGCG	5681
QY	1852	aSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuG1	1872
DB	5682	CTCCCACTCCCATGCCACACGACCTCGCCCATCTCCCTCGGACCCAGGATCCCTCCA	5741
QY	1872	nGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluPr	1892
DB	5742	GCAGAGACCCAGTGTCTTCCAAACACAGCATGAAGGGTATCATCACCGCTGTGGAGCC	5801
QY	1892	oSerLysProThrValLeuArgSerThrSerSerProValArgProAlaAlaTh	1912
DB	5802	CAGCAAGCCACCGTCTGAGGTCCACCTCCACCTCTTACCGTTGCGCCAGCTGCCAC	5861
QY	1912	rPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrrProThrLe	1932
DB	5862	ATTCCCACTGCCACCCACCTGCCCCACTGGCGCGCACCCCTCGATGGGTCTTACCTAC	5921
QY	1932	uMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProAr	1952
DB	5922	CATGGAGCCCGTCTTGTGCCAAGAGGCGCCCCCGGGTGGCCCGCAGAGCGGCGCG	5981
QY	1952	gAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAl	1972
DB	5982	AGCAGACACCGCCCATGCTTCTTCTGCCAAGCCCGCCAGCCCGCTCGGGGTGGAGCCCG	6041
QY	1972	aSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAl	1992
DB	6042	CTCCTCCCCCAGCAAGGGCTCGGAGCCCCCGGCCCTTAGTGCTCTCTGTGCGCACGC	6101
QY	1992	aThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProPr	2012
DB	6102	CACCATCGCCCGCACCCCTTCGCAAGAACCTTCGACCTCACCGCAGCCCGGACCGCC	6161
QY	2012	oAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSe	2032
DB	6162	GGCGCACCTGCTCGGCTTCGACCGCGCACCGGGAAAAGACTCAAAAGTAAACCTTTTC	6221
QY	2032	rIleGlnGluLeuGluArgSerLeuGlyTyrrHisGlySerSerTyrrSerProGluG1	2052
DB	6222	CATCCAGGAACCTCGTCTCTGTTGGTTTACCACCGGCGCAGGCTACACGCGCGAAGG	6281
QY	2052	yValGluProValSerProValSerProSerLeuThrHisAspLysGlyLeuProLy	2072

Db 1954 AGCCCGAGCTGCCTCCATGAGCTGAATGAGAGTTCTCGCTGGACAGAGAAGAAATG 2013
QY 621 GluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaAaGmet 640
Db 2014 GAACAGCCAAAGAAGGTCTCTGGAAACACGGCGCNACTGGTGGCCATCGCCCGATG 2073
QY 641 ValGlySerLysThrValSerGlnCysLysAsnPhetYrPheAsnTrpLysLysAaGln 660
Db 2074 GTGGGCTCCAAAGACTGTGTGCGAGTGTAAAGACTTCTACTTCAATACAAAGAGGCGAG 2133
QY 661 AsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaAaG 680
Db 2134 AACCTCATGAGTCTTTCAGCAGCACCAAGCTGAAGATGAGAGGAGAGAACCGCGCG 2193
QY 681 ArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProValVal 700
Db 2194 AGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAG 2253
QY 701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGlu 720
Db 2254 GAGGATGAGGAGATGAGGAGGCTCGGCGTCAAGCGGAAATGAGGAGGAGATGTGTGAGGAG 2313
QY 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla 740
Db 2314 GCTGAA-----GCC 2322
QY 741 ThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLys 760
Db 2323 ACTGTCAACACAGCTCAGACACGAGAGCATCCCTCTCTCTCACTGAGGCGGCGNAG 2382
QY 761 AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProPro 780
Db 2383 GACACAGGCGAGAATGGGCCCAAGCCCGCCAGCCACCTGGGCGCCAGCGGCGCCACCCCA 2442
QY 781 GlyProProThrProPro-ArgArgThrSerArgAlaProIleGluProThrProLase 800
Db 2443 GGGCCACCCACCCACCCAGCGAGGACATCCCGGCC-CCCACTGAGTCCACCCCGGCTC 2501
QY 800 rGluAlaThrGlyAlaProThrProProAlaProProSerProSerAlaProProPr 820
Db 2502 TGAAGCCACCTTACCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCT 2561
QY 820 oValValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGluGlu 840
Db 2562 TGTGGTCCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2621
QY 840 yGluGluGlnLysProProAlaAlaGluLeuAlaValAspThrGlyLysAlaGlu 860
Db 2622 GGAGGAGCAGAACCCCCCGGCTGAGGAGCTGGCAGTGGACACAGGGAGAGCCGAGGA 2681
QY 860 uProValLysSerGluCysThrGluGluAlaGluGluGluProAlaLysGlyLysAspAl 880
Db 2682 GCCCGTCAAGAGCGAGTGACGAGGAGNAGCCGAGGAGGCGCCGCGCAGGAGAGAGCG 2741
QY 880 aGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyLys 900
Db 2742 GGAGGCGCTGAGGCCACCGCCGAGAGGCGCTCAAGGCGAGAGAGAGAGAGAGAGAGAG 2801
QY 900 rGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerAlaTh 920
Db 2802 CGGAGGGGCCACACAGCCCAAGAGCTGGGCGGCGCCCGCCAGGAGAGGAGCTCCAGTGC 2861
QY 920 rCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerPr 940
Db 2862 CTGAGTGACAGAGGTGATGAGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2921
QY 940 oArgProSerLeuLeuThrProThrGlyAspProAlaAlaAsnAlaSerProGlnLysPr 960
Db 2922 AAGGCCAGGCTCTCAACCCGAGCTGGCGAGCCCGCGGCGCAATGTCCTCACCCAGAGGCC 2981
QY 960 oLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLy 980
Db 2982 ACTGGACCTGAAGCAGCTGAAGCAGCGAGCGGCTGCCATCCCGCCCATCCAGGTCCCAA 3041

QY 980 sValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProPr 1000
Db 3042 AGTCCATGAGCCCCCGGAGGAGCGAGCTCCACCAAGAGCGAGCTCCCGAGCCACC 3101
QY 1000 oProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProAr 1020
Db 3102 GCCACGCAAAACCTGCGAGCGGAGAGCGCCCTCAGCAGCCTGGCAGAGCCCGG 3161
QY 1020 gGlyLysSerArgSerProAlaProProAlaAspLysGlu----- 1033
Db 3162 GGCAGAGCAGAGAGCGCGCCACCCCGCCGACAGAGAGGAGAGAGAGCGCTGTGTCTT 3221
QY 1034 ----AlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProCysTrpThrSerGl 1052
Db 3222 CCAGCGCTTCGAGCGCGAGCGCCAGAGCTGCTGGGAGCCCGCTTGTGGATTCGG 3281
QY 1052 yLeuProPheProValProProArgGluValLysAlaSerProHisAlaProAspPr 1072
Db 3282 CCTGCCCTTCCCGTGCCTCCCGTGGTGTATCAAGGCTCCCGCATGCCCGGAGCC 3341
QY 1072 oSerAlaPheSerTrpAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAl 1092
Db 3342 CTCAGCGTCTCTAGCTTCCACCTGCTCCACTGCCCTGGGCTTCATGACACTGC 3401
QY 1092 aArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSerAl 1112
Db 3402 CCGGCGCTGCTCGCGCGCCACCCACCATCTCCAAACCGCTCCCTCATCTCTCTGC 3461
QY 1112 aLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGl 1132
Db 3462 CAAGCACCCAGCGCTCTCGAGAGGCAATAGGTGCCATCTCCCAAGGAATGTGCGTCCA 3521
QY 1132 nLeuHisValProTrpSerGluHisAlaLysAlaProValGlyProValThrMetGlyLe 1152
Db 3522 GTTCCACGTCCCGTACTCAGAGCATGCCAAGGCGCGGTGGGCGCTGTCCATGGGCT 3581
QY 1152 uProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLe 1172
Db 3582 GCCCTGCTCCAGTGGACCCCAAAAGCTGCGACCTTCAGCGGAGTGAAGCAGGAGCT 3641
QY 1172 uSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAl 1192
Db 3642 GTCCCGCAGCGGCGCAGGCTGGGCGCACCGGAGAGCCTGGGCGGTGCCACAGCCAGAGGC 3701
QY 1192 aSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyI 1212
Db 3702 GTCCGTGCTGAGAGGAGCAGCTCTGGGCTCAGTTCCGGCGGAGAGCATCACCAAGGCAT 3761
QY 1212 eProSerThrArgValProSerAspSerAlaIleThrTrpArgGlySerIleThrHisGl 1232
Db 3762 TCCAGCACACGGGTGCTCTCGACAGCGCCATCATACCGCGGCTCCATCACCCACCG 3821
QY 1232 yThrProAlaAspValLeuTrpLysGlyThrIleThrArgIleIleGlyGluAspSerPr 1252
Db 3822 CACGCGAGTGCCTGCTTACAAAGGCGCACCATCACAGGATCATCGGCGAGGAGAGCC 3881
QY 1252 oSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTrpGluGl 1272
Db 3882 GAGTCTGCTGGACCGCGCGCGGAGGAGCAGCTGCCCAAGGGCCAGCTCATCTACGAGG 3941
QY 1272 yLysLysGlyHisValLeuSerTrpGluGlyMetSerValThrGlnCysSerLysGl 1292
Db 3942 CAAGAAGGCCACGCTTGTCTATGAGGCTGGCATGTCTGTGACCCAGTGTCTCCAGGA 4001
QY 1292 uAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThrTrpAs 1312
Db 4002 GGACGCGAGAGCAGCTCAGGACCCCGCCCATGAGAGCGCGCGCCCGCCCAAGCGACCTATGA 4061
QY 1312 pMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGl 1332
Db 4062 CATGATGAGGCGCGCGTGGGCGAGGCCATCTCTCAGCGAGCATCGAAGGTCTCATGG 4121

SHSHAHQHSPISPRTQDALQORPSVLNHTCMKGIITAVEPSKPTVLSTSTSSVRPA
ATFPPTHCPGLGTLGVPTLMPEVLLPKFAPRVARPERPRADTGHAFAPPARSG
LEPASPSKSEPRPLVPVSHATARTPAKNLPHASPDPPAPPASADPHREKT
QSKPFSQELERSLGVHGSYSPEGVSPVSPSLTHDKLPHUEELDKSHLEG
ELAPKQPVKLGEEAHLPHLPLPESOPSSPLLOTAPGVKHQHVVTTLAQHISEV
ITODYTRHPHQOLSAPLPAPLYSPFCASCPVLDRPPSDLYLPPPDHGAARGSPHS
EGKRSPFNKTSVLGGEDGIBPVSPPEGMTEPHGRSAVPLLYRDGEOTEPRMG
SKSPGNTSQPPAFKLITSENANWVKQBEINKKLNKUNTRNEPEINISQPGTEIFNM
PAITGTGLMYTRSQVQVEHASTNMGLEAIIRKALMKQIDOWEESPISLANAFLNAS
ASLPAAMPITAAQDRSDHITLTPSGGCKAKVSPSSRKAKGAPAGLASGDRPPSVSS
VHSEGCNRRTPLTNRWEDRPSAGSTPPFYNPLIMRLQAGVMASPPPPGLPAGSGP
LAGPHHAWDEEPKLLCSQVETILSDSE"

ORIGIN

Alignment Scores:
Pred. No.: 1,26e-254 Length: 8686
Score: 12978.00 Matches: 2487
Percent Similarity: 98.57% Conservative: 3
Best Local Similarity: 98.46% Mismatches: 9
Query Match: 98.21% Indels: 28
DB: Gaps: 3

US-09-522-753-5 (1-2517) x AF125672 (1-8686)

Qy	1	MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro	20
Db	157	ATGTGCGGCTCCACACAGCCTGTGGCACAGCTGGAGGGCCACTGAGCCCGCTACCCG	216
Qy	21	ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu	40
Db	217	CCCCACAGCCTTCTACCCAGTCAGATCCCGCGGACGACACGCGAGCTCGGCTCTG	276
Qy	41	GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln	60
Db	277	GAGTACCAGCACCACTCCCGGAGCTATGCTCCACCTCGCCGCGCTCCATCATCCAG	336
Qy	61	ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln	80
Db	337	CCCCAGCGCGAGCGCTCCCTGCTGCTGAGTTCAGCTCCAGCCCGGAGTGAACGCTCCAG	396
Qy	81	GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet	100
Db	397	GAGCTCCACCTGCGGCCAGAGTCCCACTCATACCTGCCGAGCTGGGGAAATCAGAGATG	456
Qy	101	GluPheIleGluSerLysArgProArgGluLeuLeuLeuProAspProLeuLeuArgPro	120
Db	457	GAGTTTATTGAAGCAAGCGCCTCGGTGAGTGTGCTGAGCTGCTGACCCCTCTCGGACCG	516
Qy	121	SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer	140
Db	517	TCACCCCTGTGCCCACCGGCCAGCTCGGGATCTGAAGACCTCACCAAGGACCGTAGC	576
Qy	141	LeuThrGlyLysLeuGluProValSerProProSerProHisThrAspProGluLeu	160
Db	577	CTGACGGGCAAGCTGGAACCGGTGTCTCCCGCCAGCCCGCCGACACTGACCCCTGAGCTG	636
Qy	161	GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp	180
Db	637	GAGCTGGTGGCCCGCCAGCGCTGTCGAAGGAGGAGCTGATCCAGAACATGGACCGGTGGAC	696
Qy	181	ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeu	200
Db	697	CGAGAGATCACCATGTTGAGCAGCAGATCTCTAAGCTGAAGAGAAGCAACAGCAGCTG	756
Qy	201	GluGluGluAlaAlaLysProGluProGluLysProValSerProProIleGlu	220
Db	757	GAGGAGGAGGCTGCCAAGCGCCGAGCTGAGAGCGCGTGTACCCCGCCCATTCGAG	816
Qy	221	SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAla	240
Db	817	TGAAGCACCGGAGCTGTGTGAGATCATCTACGACGAGACCCGGAAGAGGCTGAAGCT	876
Qy	241	AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro	260

Db	877	GCACATCGGATTCTGGAAGGCTCGGGGCCCCAGGTGGAGCTGCGCTGTGTACACGAGCCC	936
Qy	261	SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu	280
Db	937	TCCGACACCCGGCAGTATCATGAGAACATCAAAATAAACACAGGCGATCGGAGAGACTA	996
Qy	281	IleLeuTyrPheLysArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGln	300
Db	997	ATCTTGTTACTTCAAGAGGAGGAAATCACGCTCGGAAACAATGGGACGACAGTCTTGCCAG	1056
Qy	301	ArgTyrAspGlnLeuMetGluAlaLeuLysLysValGluArgIleGluAsnAsnPro	320
Db	1057	CGCTATGACCAGCTCATGGAGGCTTGGGAGAAGAGTGGAGCGCATCGAACAACCCC	1116
Qy	321	ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle	340
Db	1117	CGGCGCGCGGCCCAAGGAGAGCAAGGTTCCGAGTACTACGAGAGCAGTTCCTTGAGATC	1176
Qy	341	ArgLysGlnArgGluLeuGlnArgMetGlnSerArgValGlyGlnArgGlySerGly	360
Db	1177	CGCAAGCAGCGCGAGCTGCAGGAGCGCATGCAG---AGGGTGGGCGCAGCGGCGCAGTGGG	1233
Qy	361	LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer	380
Db	1234	CTGTCCATGTCGCCCGCCGCGCAGCAGCAGGAGTGTGAGAGATCATCGATGGCTCTCA	1293
Qy	381	GluGlnGluAsnLeuLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr	400
Db	1294	GAGCAGGAGAACTCGAGAGCAGATGCGCCAGCTGGCGTGTATCCCGCCCATGCTGTAC	1353
Qy	401	AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet	420
Db	1354	GAGCTGACCCAGCAGCGCATCAAGTTTCATCAACATGAACGGCTTATGTCGCCACCCATG	1413
Qy	421	LysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGlyLysGluThrPhe	440
Db	1414	AAGGTGTACAAAGACCGCCAGGTCTAGAACATGTGAGGTGAGCAGAGAGAGACCTTC	1473
Qy	441	ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg	460
Db	1474	CGGAGAGAGTTCATGAGCATCCCAAGACTTTGGCTGATCGCATTCCTCTGGAGAGG	1533
Qy	461	LysThrValAlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyrLys	480
Db	1534	AACACAGTGGCTGAGTGGCTCTCTATTACTACTGACTGACTAAGAAGAATCAGAACTAAG	1593
Qy	481	SerLeuValArgArgSerTyrArgArgGlyLysSerGlnGlnGlnGlnGlnGln	500
Db	1594	AGCTTGTGAGCGGAGCTATCGGCCCGCGGCAAGAGCCAGCAGCAACACAGCAGCAG	1653
Qy	501	GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	520
Db	1654	CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	1713
Qy	521	AspGluLysGluLysGluAlaGluLysGluGluLysProGluValGluAsn	540
Db	1714	GATGAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1773
Qy	541	AspLysGluAspLeuLeuLysThrAspThrAspThrSerGlyGluAspAsnAspGlu	560
Db	1774	GACAGGAACCTCTCTCAGGAGAGAGACGACACACCTCAGGGGAGGAGCAGCAG	1833
Qy	561	LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGly	580
Db	1834	AAGGAGGCTGTGGCTCCAAAGGCGCAAACTCCCAACAGCCAGGAGGAGGAGGAGGAG	1893
Qy	581	ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln	600
Db	1894	CGCATACCCCGCTCAATGGCTAATGAGGCCAACAGCGAGGAGGCCATCACCCCGCAG	1953
Qy	601	SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluMet	620

Db 4262 GCGGGCGCTCCATCCATGAGATCCCGCGGAGGAGCTCGGGACACGCGCCGAGCTGCC 4321
Qy 1441 LeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAsp 1460
Db 4322 CTGGCCCGCGCGCTCAAGGAGGCTCCATCAGCAGGCGACCCGCTCAAGTACGAC 4381
Qy 1461 ThrGlyAlaSerThrGlySerLysLysHisAspValArgSerLeuIleGlySerPro 1480
Db 4382 ACCGGCGCTCCACCACTGGCTCCAAAAGACGACGATGAGTCCCTCATCGCGAGCC 4441
Qy 1481 GlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGlu 1500
Db 4442 GCGCGAGGTTCACCCCGTGACCCCTGATGTGATGCGCCAGCGCGGCACTGGAA 4501
Qy 1501 ArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGly 1520
Db 4502 CGTCCCTGCTACGAGGAGGCTGAGAGCGCGCCAGGACCGCCAGCAGCTCGGGGGC 4561
Qy 1521 SerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerPro 1540
Db 4562 TCCATTGCGCGCGCGCGCTGATTTGTGCTGAGCTGGGTAAAGCGCGGAGCC 4621
Qy 1541 LeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProVal 1560
Db 4622 CTGACCTATGAGGACACCGGGGACCTTTTGGCGGCCACCTCCACGAGGTTCGCCGTG 4681
Qy 1561 ThrMetArgGluProThrProArgLeuGlnGlySerLeuSerSerSerLysAlaSer 1580
Db 4682 ACCATGCGGGAGCCACGCGCGCTCGCAGGAGGCGAGCTTTGCTCCACGAGGATCC 4741
Qy 1581 GlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrVal 1600
Db 4742 CAGGACCGAAGCTGACGCTGACGCGCTCGTGAATGCCAAGTCCCGCAGCAGCCGTG 4801
Qy 1601 ProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerGly 1620
Db 4802 CCGGAGCACCCACACCCCATCTCGCCCTATGAGCACCTGCTTCGGGGCGTGAGTGGC 4861
Qy 1621 ValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGly 1640
Db 4862 GTGACCTGTATCGCAGCACATCCCTCGCTTCGACCCCATCTCCATACCCCGCGGC 4921
Qy 1641 IleProLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThr 1660
Db 4922 ATCCCTCTGAGCGAGCGCTGCTGCTACTCTGCCCCGACACTTGGCCCCCAACCCACC 4981
Qy 1661 TyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGlu 1680
Db 4982 TACCCGACCTGTACCCACCTTACCTCATCCGCGCTACCCCGACACGCGCGCTGGAG 5041
Qy 1681 AsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAla 1700
Db 5042 AACCCGCGAGACCATCATCAATGACTACATCACCTCGCAGCAGATGCCACCAACAGGCC 5101
Qy 1701 ThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeu 1720
Db 5102 ACCGCCATGCGCCAGCGAGCTGATATGCTGAGGGGCTCTGCGCCCGCGAGTCTCTCGCTG 5161
Qy 1721 AlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeu 1740
Db 5162 GCACCTCACTAGCTGCGGGTCCCGAGGCATCATCGACCTGTCCAAAGTGCACACCTG 5221
Qy 1741 ProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyr 1760
Db 5222 CTTGTGCTGTGCCCCCGACACGAGCACCCAGCCACCGCCATGAGCCGCTTGCCTTAC 5281
Qy 1761 LeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGly 1780
Db 5282 CTCCCCACCGCGCCCGAGCCCTTCAGCAGCGGCCACAGCAGCTCCCTCCACGAG 5341
Qy 1781 GlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArg 1800
Db 5342 GGTCCAACACACTTGAAAAACCAACCAACCATCTCTCTCGTCCGAGCGGAGCGAGCCGG 5401

Qy 1801 AspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThr 1820
Db 5402 GATCGAGAGCGGAGCCGGATCGGAGCGGGGAAAGTCCATCTCTCATCGTCCACCGACG 5461
Qy 1821 ValGluHisAlaProIleTyrArgProGlyThrGluGlnSerSerGlySerSerGlySer 1840
Db 5462 GTGAGAGCAGCAGCCATCTGGAGACCTGTGTACAGAGCAGAGCAGCGCAGCGCGGAGC 5521
Qy 1841 SerGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis 1860
Db 5522 AGCGCGGGGTGGGGGAGCAGCAGCGCCCGCCCTCCACTCCATGCCATGCCACCGAC 5581
Qy 1861 SerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsn 1880
Db 5582 TCGCCCATCTCCCTCGGACCCAGGATGCCCTCCAGCAGAGCAGCAGTGTGTTCACAAC 5641
Qy 1881 ThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArgSer 1900
Db 5642 ACAGGCATGAAGGGTATCATCACCGCTGTGGAGCCAGCAAGCCACCGTCTGAGGTCC 5701
Qy 1901 ThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysPro 1920
Db 5702 ACCTCCACCTCTCACCCGTTGCGCCAGCTGCCCATTTCCACCTGCCACCCACTGCCCA 5761
Qy 1921 LeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLys 1940
Db 5762 CTGGGCGGACCCCTCGATGGGTCTTACCCTTACCTCATGAGCGCGCTTCTGTGCCAAG 5821
Qy 1941 GluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeu 1960
Db 5822 GAGGCCCCCGGGTCCCGCGCAGAGCGGCCCCGAGCAGACACCGGCCATGCTTCTCTC 5881
Qy 1961 AlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGlu 1980
Db 5882 GCGAAGCCCCGAGCGCTCGGGCTGGAGCCCGCTCTCTCCCCAGCAAGGGCTCGGAG 5941
Qy 1981 ProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProLys 2000
Db 5942 CCGCGCCCTAGTGTCTCTCTGTGGCCACGCCATCGCCCGCACCCCTCGGAG 6001
Qy 2001 AsnLeuAlaProHisHisAlaSerProAspProProAlaProAlaSerAlaSerAsp 2020
Db 6002 AACCTCGCACCTCACCGCCAGCCCGACCCGCGCGGCCACTTGCCTCGGCTCGGAC 6061
Qy 2021 ProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuIleArgSer 2040
Db 6062 CCGCACCGGGAAAGACTCAAAGTAAACCTTTTCCATCCAGGAACTGGAACTCCGTCTCT 6121
Qy 2041 LeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSer 2060
Db 6122 CTGGGTATCCACCGCAGCAGCTACAGCCCCGAAGGGGTGGAGCCCTCAGCCCTGTGAGC 6181
Qy 2061 SerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluLeuLeuAspLysSer 2080
Db 6182 TCACCCAGTCTGACCCACGACGAGGGCTCCCCAACACACTTGGAGAGCTCGAAGAGC 6241
Qy 2081 HisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAla 2100
Db 6242 CACTCGAGGGGAGCTGCGGCCCAAGCAGCAGCGCCCGGTGAAGCTTGGCGGGAGGCC 6301
Qy 2101 AlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuLeu 2120
Db 6302 GCCACCTCCACACCTTGGCGCGCTGCTGAGAGCAGCCCTTCTGTCCAGCCCGCTGCTC 6361
Qy 2121 GlnThrAlaProGlyValLysGlyHisGlnArgValThrLeuAlaGlnHisIleSer 2140
Db 6362 CAGACCGCCCCAGGGTCAAAGGTCAACAGGGGTGCTACAGCGGTGTCACTGCGCCACATCAGT 6421
Qy 2141 GluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuPro 2160
Db 6422 GAGGTCTATCACAGGACTACACCCGGCACCACCCAGCAGCTCAGCGCACCCCTTGCCC 6481

QY 701 GluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGlu 720
DB 2102 GAGCATGAGGAGTGGAGGCGTCGGCGTTCAGCGGAAATGAGGAGGAGATGGTGGAGGAG 2161
QY 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla 740
DB 2162 GCTGAAGCCCTTACATGCTCTGGGAATGAGTCCAGAGGGGAATGCACTGGCCACGCC 2221
QY 741 ThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLys 760
DB 2222 ACTGTCAACACAGCTCAGACACCGAGAGCATCCCTCTCTCCTCACACTGAGGCGGCCAG 2281
QY 761 AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProPro 780
DB 2282 GACACAGGGCAGAATGGGCCCAAGCCCGCCAGCCACCTGGGGCGCCAGCGGCCACCCCA 2341
QY 781 GlyProProThrProProArgArgThrSerArgAlaProIleGluProThrProAlaSer 800
DB 2342 GGCCCCACCCACCCACACGAGGAGCATCCCGGGCCCCCATTCAGGCCACCCCGGCTCT 2401
QY 801 GluAlaThrGlyAlaProThrProProAlaProProSerProSerAlaProProPro 820
DB 2402 GAAGCCACCGAGCCCTACGCCCCACACGACGCCATCGCCCTCTGCACCTCTCTCT 2461
QY 821 ValValProLysGluGluLysGluGluThrAlaAlaAlaProProValGluGluGly 840
DB 2462 GTGGTCCCCAAGGAGGAGGAGGAGGAGACCGCAGCAGCGCCCCCAGTGGAGGAGGG 2521
QY 841 GluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGlu 860
DB 2522 GAGGAGCAGAAGCCCCCGCGCTGAGAGCTGGCAGTGGACACAGGGAAGGCCGAGGAG 2581
QY 861 ProValLysSerGluCysThrGluGluAlaGluGlyProAlaLysGlyLysAspAla 880
DB 2582 CCGCTCAAGACGAGTGACGAGGAGAACGCCAGAGGGGGCCGCCAAGGGCAGGAGCG 2641
QY 881 GluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySer 900
DB 2642 GAGGCCGCTGAGGCCACCGCGGAGGGGGCGCTCAAGCAGAGAAGAAGGAGGGCGGAGC 2701
QY 901 GlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerAlaThr 920
DB 2702 GGAGGGGCCACCTGCTCAAGAGCTGGGGCGCCCCCAGGACAGCACTCCAGTGTACC 2761
QY 921 CysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerPro 940
DB 2762 TGCAGTGCAGACGAGGTGGATGAGCGCAGGGCGGCGACAGAACCGGCTGTCTCCCA 2821
QY 941 ArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysPro 960
DB 2822 AGGCCAGCCCTCTCACCCCGACTGGCGACCCCCGGGCCAATGCCCTCACCCCAAGAGCCA 2881
QY 961 LeuAspLysGlnLeuLysGlnArgAlaAlaIleProProIleGlnValThrLys 980
DB 2882 CTGGACCTGAAGCAGCTGAAGCAGCAGCGGCTGCCATCCCCCCCATCCAGGTCAACAA 2941
QY 981 ValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProPro 1000
DB 2942 GTCAATGAGCCCCCGGAGGAGCGAGCTCCACCAAGCCAGCTCCCCCAGGCCACCG 3001
QY 1001 ProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArg 1020
DB 3002 CCACCGCAAAACCTGCAGCGCGGAGAGCGACGCCCTCAGCAGCTGGCAGCAGCCCCCG 3061
QY 1021 GlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGln 1040
DB 3062 GGCAGAGCAGGAGCGCGGACCCCGCCGCAAGAGAGCCCTTCGCGAGCGCGAGGCCAG 3121
QY 1041 LysLeuProGlyAspProProCysTTPThrSerGlyLeuProPheProValProProArg 1060
DB 3122 AAGCTGCTGGGAGACCCCTTGTGGACCTTCGGGCTGGCTTCCCGTGGCCCCCGCT 3181
QY 1061 GluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyAlaProPro 1080

DB 3182 GAGGTGATCAAGGCGCTCCCGCATGCCCGAGCCCTCAGCCCTTCTCTACGCTCCACCT 3241
QY 1081 GlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProPro 1100
DB 3242 GGTACCCACTGCCCTTGGGCTTCATGACACTGCCGCGCGCTCTCTCGGCGGCCACCC 3301
QY 1101 ThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuGluArg 1120
DB 3302 ACCATCTCAACCCCGCTCCCTCTCTCTCTGTCGCAAGCACCCAGCGCTCTCGAGAGG 3361
QY 1121 GlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyTrSerGluHis 1140
DB 3362 CAATAGTGTGCATCTCCCAAGGAATGTCTGGTCCAGCTCCAGTCCCGTACTCAGAGCAT 3421
QY 1141 AlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLys 1160
DB 3422 GCALAGGCCCGGTGGGCTCTGTACCATGGGGCTGCCCTGCCATGAGCCCCCAAAAG 3481
QY 1161 LeuAlaProPheSerGlyValLysGlnGluLeuSerProArgGlyGlnAlaGlyPro 1180
DB 3482 CTGGCACCCCTTCAGCGGAGTGAAGCAGGAGCAGCTGTCCCCACGGGCCAGGCTGGGCCA 3541
QY 1181 ProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeu 1200
DB 3542 CCGGAGAGCTGGGGTGGCCACAGCCCAAGGAGCGTCCGCTGAGAGGAGCAGCTCTG 3601
QY 1201 GlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAsp 1220
DB 3602 GGTCTAGTTCCGGGCGGAAGCATCAACAAAGCATTCACAGCACCGGTGCCCTCGGAC 3661
QY 1221 SerAlaIleThrTyArgGlySerIleThrHisGlyThrProAlaAspValLeuTyLys 1240
DB 3662 AGCGCCATCACATACCGCGCTCCATCACCCACGGCAGCCAGCTCAGCTCTGTACAAG 3721
QY 1241 GlyThrIleThrArgIleIleGlyLeuAspSerProSerArgLeuAspArgGlyArgGlu 1260
DB 3722 GGCACCATCACCAGGATCATCGCGAGGACAGCCCGAGTCTGGACCGCGCGCGGAG 3781
QY 1261 AspSerLeuProLysGlyHisValIleTyGlyGlyLysGlyHisValLeuSerTyLys 1280
DB 3782 GACAGCTGCCACAGGCCACCGTCTCTACGAAGCAAGAGGCCACGCTCTTCTCTAT 3841
QY 1281 GluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerGlyPro 1300
DB 3842 GAGGTGGCATGTCTGTGACCCAGTCTCCAAGGAGACGGCAGAGCAGCTCAGGACCC 3901
QY 1301 ProHisGluThrAlaAlaProLysArgThrTyArgMetMetGluGlyArgValGlyArg 1320
DB 3902 CCCCATGAGAGCGCGCCCAAGCGCCTATGACATGATGAGGGCGCGCGTGGGCGAG 3961
QY 1321 AlaIleSerSerAlaSerIleGlyLeuMetGlyArgAlaIleProProGluArgHis 1340
DB 3962 GCATCTCTCAGCCAGCATCGAGGTCTATGGGCGCTGCCATCCCGCGGAGCGACAC 4021
QY 1341 SerProHisIleLysGlnGlnHisIleIleArgGlySerIleThrGlnGlyIlePro 1360
DB 4022 AGCCCCACCCCTCAAGAGCAGCAGCACCATCCCGGGTCCATCACAAAGGATCCCT 4081
QY 1361 ArgSerTyValGluAlaGlnGluAspTyLeuArgArgGluAlaLysLeuLysArg 1380
DB 4082 CGGTCTAGCTGGAGGACAGGAGGACTACCTGCTCGGAGGCCAAGCTCTCTAAAGCGG 4141
QY 1381 GluGlyThrProProProProProSerArgAspLeuThrGluAlaTyLysThrGln 1400
DB 4142 GAGGGCAGCCCTCCCGCCCCCAGCCCTCAGCGGACCTCAGCGAGCCCTACAGAGCGAG 4201
QY 1401 AlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGlu 1420
DB 4202 GCCCTGGGCCCTTGAAGCTGAAGCGGCCCATGAGGGCTGGTGGCCACGGTGAAGGAG 4261
QY 1421 AlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuPro 1440

Alignment scores:
Pred. No.:
Score:
Percent Similarity:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 15, 2004, 22:23:25 ; Search time 20448 Seconds
(without alignments)
5335.212 Million cell updates/sec

Title: US-09-522-753-5
Perfect score: 13215
Sequence: 1 MSGSTQLVAQTWRATEPRYP.....WDEPKPLCSQVETILSDSE 2517

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2 1/USPTO.spool/US09522753/runat_15042004_143737_17428/app_query.fasta_1.2695
-DB=GenEmbl -OPMT=fastap -SUFFIX=xge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=100
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=100 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09522753 @CGN 1_14601@runat_15042004_143737_17428 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_ov.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_hum.*
19: em_mu.*
20: em_mu.*
21: em_mu.*
22: em_mu.*
23: em_mu.*
24: em_mu.*
25: em_mu.*
26: em_mu.*
27: em_mu.*
28: em_mu.*

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	13215	100.0	8561	9	AF113003	AF113003 Homo sapi
2	12978	98.2	8686	9	AF1125672	AF1125672 Homo sapi
3	11067	83.7	8388	10	AF113001	AF113001 Mus muscu
4	10832	82.0	8544	10	AF125671	AF125671 Mus muscu
5	9894	74.9	7465	10	AF113002	AF113002 Mus muscu
6	7885	59.7	5989	9	HSU37146	U37146 Human silen
7	4612	34.9	2930	9	S83390	S83390 T3 receptor
8	4194	31.7	7949	9	AB028970	AB028970 Homo sapi
9	4187	31.7	7780	10	MMU35312	U35312 Mus musculu
10	4147	31.4	7940	6	AX578069	AX578069 Sequence
11	4147	31.4	7940	9	AF044209	AF044209 Homo sapi
12	4043	30.6	8018	9	AF087856	AF087856 Homo sapi
13	4009	30.3	8959	5	AF495886	AF495886 Xenopus 1
C 14	3736	28.3	194840	2	AC139377	AC139377 Mus muscu
C 15	3479	26.3	254449	2	AC097560	AC097560 Rattus no
C 16	3445	26.1	218129	2	AC136560	AC136560 Rattus no
17	3344	25.3	219339	2	AC121005	AC121005 Rattus no
18	3178	24.0	2842	9	BC004326	BC004326 Mus muscu
19	3012	22.8	2964	10	BC047524	BC047524 Mus muscu
20	2891	21.9	205283	9	AF073916	AF073916 Homo sapi
C 21	2187	16.5	3120	9	AF030586	AF030586 Homo sapi
22	2155	16.3	1917	5	BC054296	BC054296 Xenopus 1
23	2091	15.8	6328	10	AB093281	AB093281 Mus muscu
24	1900	14.4	6541	9	AB019524	AB019524 Homo sapi
25	1839	13.9	4285	10	MUSRI13G	L78294 Mus musculu
26	1687	12.8	1891	9	BC050594	BC050594 Homo sapi
27	1663	12.6	1850	9	BC058511	BC058511 Homo sapi
28	1633	12.4	1741	9	BC056862	BC056862 Homo sapi
29	1628	12.3	3025	5	BC049302	BC049302 Danio rer
30	1396	10.6	161970	2	AC027706	AC027706 Homo sapi
31	1391	10.5	2914	9	AK127788	AK127788 Homo sapi
32	1272	9.6	3575	9	HSB02410	AL137641 Homo sapi
33	1172	8.9	3997	9	AF030585	AF030585 Homo sapi
C 34	1162	8.8	133947	5	AL590153	AL590153 Zebrafish
35	1113	8.4	650	6	AF175223	AF175223 Sequence
36	1099	8.3	11296	3	AF175223	AF175223 Drosophil
37	1076	8.1	872	6	BD21548	BD21548 Human gen
38	960	7.3	555	6	AX677866	AX677866 Sequence
39	932	7.1	560	9	HSU80750	U80750 Homo sapien
C 40	932	7.1	560	9	HSU80761	U80761 Homo sapien
C 41	885	6.7	718	6	AX753058	AX753058 Sequence
C 42	845	6.4	710	6	BD021376	BD021376 Novel gen
C 43	845	6.4	710	6	BD101314	BD101314 Novel gen
44	774	5.9	673	6	AX677865	AX677865 Sequence
45	763	5.8	135351	9	AC068279	AC068279 Homo sapi
46	760	5.8	150769	2	AC130376	AC130376 Homo sapi
47	750	5.7	1793	9	AK057740	AK057740 Homo sapi
48	746	5.6	520	6	AX396270	AX396270 Sequence
C 49	743	5.6	40871	2	AC020019	AC020019 Drosophil
C 50	743	5.6	160440	3	AC023741	AC023741 Drosophil

QY 2072 LysHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnPro 2091
DB 1563 AAGCACTGGAGAGCTCGACAAGAGCCACTCGAGGGGAGCTGGGCCCAAGCAGCCA 1622
QY 2092 GlyProValLysLeuGlyGlyGluAlaHisLeuProHisLeuArgProLeuProGlu 2111
DB 1623 GGCCCCGTGAAGCTTGCGCGGGAGCGGCCCACTCCACACCTGGCGCGCTGCCTGAG 1682
QY 2112 SerGlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArg 2131
DB 1683 AGCAGCCCTCGTCCAGCCCGCTGCTCCAGACCGCCCGAGGGGTCAAGGTCAACGCG 1742
QY 2132 ValValThrLeuAlaGlnHisLeuSerGluValLeuThrGlnAspTyrThrArgHisHis 2151
DB 1743 GTGGTCACTGGCCAGCAGCATCAGTGGGTCAATCACACGAGACTACACCCGCGCAC 1802
QY 2152 ProGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCys 2171
DB 1803 CCACAGAGCTCAGCGACCCCTGCGCGCCCTCTACTCTTCCCTGGGGCCAGCTGC 1862
QY 2172 ProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProAspHisGly 2191
DB 1863 CCGTCTCTGAGCTTCCGCCGCCACCCAGTGAATCTTACCTCCGCCCGCGGACCATG 1922
QY 2192 AlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLys 2211
DB 1923 GCCCGCGCGTGGCTCCCCCAGCAGAAAGGGGCAAGAGTCTCCAGAGCCAAACAAG 1982
QY 2212 ThrSerValLeuGlyGlyGluAspGlyLeuProValSerProProGluGlyMet 2231
DB 1983 ACCTCGCTCTTGGGTGGTGGAGCGGTATTTGAACCTGTCTCCACCGAGGGCGATG 2042
QY 2232 ThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGln 2251
DB 2043 ACCGAGCCAGGCACTCCCGAGTGTGTGTACTCCCGCTGTACTCCGGATGGGGACAG 2102
QY 2252 ThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhe 2271
DB 2103 ACCGAGCCAGCAGGATGGCTCCAGTCTCCAGGCAACACAGCAGCGCCGCGCTTC 2162
QY 2272 PheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysGlnGlnLeuAsn 2291
DB 2163 TTCAAGCAAGCTGACCGAGACCACTCCGCCATGGTCAAGTCCCAAGAACGAGATCAAC 2222
QY 2292 LysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnLeuSerGlnProGlyThr 2311
DB 2223 AAGAGCTGAACCAACCAACCGGATGAGCTGATACATATCAGCAGCTGGGAGC 2282
QY 2312 GluLeuPheAsnMetProAlaLeuThrGlyThrGlyLeuMetThrTyrArgSerGlnAla 2331
DB 2283 GAGATCTTCAATATGCCCGCATCACCGGAACAGGCTTATGACCTATAGAACGAGCGC 2342
QY 2332 ValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaLeuLeuLeuArgLysAlaLeuMet 2351
DB 2343 GTGAGGAACATGCCAGCAACATGGGCTGGAGGCCATAATTAGAAAGGCACCTCATG 2402
QY 2352 GlyLysTyrAspGlnTyrGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeu 2371
DB 2402 ----- 2402
QY 2372 AsnAlaSerAlaSerLeuProAlaAlaMetProLeuThrAlaAlaAspGlyArgSerAsp 2391
DB 2402 ----- 2402
QY 2392 HisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyArgProSerSer 2411
DB 2403 -----GGTGGCGCGGGAAGGCCAAGGTCTCTGGCAGACCCAGCAGC 2444
QY 2412 ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSer 2431
DB 2445 CGAAAGCCAAAGTCCCCCGGCCCGGCTGGATCTGGGACCGGCCACCTCTGCTCC 2504
QY 2432 SerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTyrGlu 2451

DB 2505 TCAGTGCACTCGAGGAGGAGACTGCAACCCCGGACCGCTACCAACCGGTGGGAG 2564
QY 2452 AspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuLeuMetArgLeu 2471
DB 2565 GACAGGCCCTCGTCCGAGGTTCCAGCCATCCCTACACCCCTGATCATGGGCTG 2624
QY 2472 GlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeu 2491
DB 2625 CAGCGGGTGTATGCTTCCACCCCGGCTCCCGCGGCGAGCGGCGCCCTC 2684
QY 2492 AlaGlyProHisHisAlaTrpAspGluProLysProLeuLeuCysSerGlnTyrGlu 2511
DB 2685 GCTGGCGCCACACACCGCTGGGACGAGGCCAAGCCACTGCTCTGCTGCAGTAGCAG 2744
QY 2512 ThrLeuSerAspSerGlu 2517
DB 2745 ACACTCTCCGACAGCAG 2762
RESULT 8
AB028970 7949 bp mRNA linear PRI 02-APR-2003
LOCUS Homo sapiens mRNA for KIAA1047 protein, partial cds.
DEFINITION AB028970
ACCESSION AB028970
VERSION AB028970.2 GI:29421187
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Kikuno,R., Nagase,T., Ishikawa,K., Hirotsawa,M., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro
JOURNAL DNA Res. 6 (3), 197-205 (1999)
MEDLINE 99397452
PUBMED 10470851
REFERENCE 2 (bases 1 to 7949)
AUTHORS Ohara,O., Nagase,T. and Kikuno,R.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1999) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail:cdmainfo@kazusa.or.jp. Tel: +81-438-52-3913, Fax: +81-438-52-3914)
COMMENT On Mar 31, 2003 this sequence version replaced gi:5689430.
FEATURES
Location/Qualifiers
1..7949
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="17"
/clone="hh0122181"
/tissue_type="brain"
/notes="vector: pBC SK plus"
1..7949
/gene="KIAA1047"
/contig="7058"
/notes="KIAA1047"
/notes="Start codon is not identified.
This sequence was obtained by subcloning of the DNA fragments derived from two cDNA clones (1 - 1691 was derived from hh01221 and 1692 - 7949 was derived from hg02672)."
codon_start=1
/product="KIAA1047 protein"
/protein_id="BAA82999.2"
/db_xref="GI:29421188"
/translation="SHSLVTRSLIMSSGYPNPGAFSTEQSRYPHPSVQYTFPNTR HQGEFAYVPYRSHLVSQASQLQQQQQLRRPRLSEFHPGSDRPOERTSYEP FHPGPSVDHLSLEKPRLEQVDSHFQVSAVPLVPLHPLFGLRASADAKDPAP GKGHEAPSPISGQPCGDDQNASPSKLSKEELIQSDMRDVRDREIAKVEQILKLKKQO

QLEEEAAKPEPEKPVSPPEVEQKHSIVQIYYDENRKAEEAHKIFEGLGKVELPL
 YNOSPSTKVYHNIKTQVNRKKLILFFKRKHARKOREQKICORVDQSLMEAWKKVDS
 RINNNRRKAKESKTEYREYKQPEIRKOREQOERFORVGARGAQSLATSEHSEHIS
 EIIDGLSEQENNEKQMRQLSVIPPMFDEAQRVKELINWGLMEDPMKYKQRPVNV
 WTDHEKEIFADKPIQKPGFLIASLERSKSPDCVLYIYLTIKNENIKALVRNKGK
 RRGNQOIAQPEEKVEKBEKDEKKEKKEKBEKKEKDEKKEKDEKKEKDEKID
 GTABETEERQATPRKGTANSQRRKGRITRSMTNEAAAAAATAEPPPLPP
 PPEPISTPEVTSRWTEEMEVAKGLVEHGRNAAIAKVVGTGKSAQCKNPFVNYKR
 RHNLDLLOQHKOKTSRKRREERDVOCESVASTVGAODEDIEASNEENPEDSGA
 KNSDNLQVAVKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEK
 TKPAEDSVETQVNDISIAETAQMDVDOQESHAERGSVCDPPPKATKADSVDVEVRVP
 ENHSAVEGNTKMDLDRASEKVEPRDLDVVAQINQAPPEPOSDNDSSATCSADE
 DVNDEGERQRMFEMDKSLNPTGSLVSSLPKNPLDLPOLQHRAAVIPPVMSCTP
 CNLIPITPQVAFYALYOHKIKAMESALLBEORQORQIDLECRSSPCGTSGKSPNRE
 WEVLQAPHQVITNLPEGVRLPTTRPPRPLIPSSKTTVAESKPSFTMGGSISQGT
 PCGYLTSHNOASVTOETPSPVSGISLGLPQOESAKSATLPYKQEEFSPRSQNSQP
 EBLVRAQHEGVVRGTAGALQEGSITRGTPTSKISVESIPSLSGSITQGTPTPALPTQGI
 RTHAIEISLRKSYSEVGNIKQMSRESPVAPLEGLICRALPGRSPHSDLKERTVLS
 GSIQGTPTATTESFEDGLYKPKQIKRESPPPIRAFEAGITKGPYDGIITIKEMGRSI
 HEIPRODILTOBSRKTPEVVQSRPIIEGSIQGTPIKFDNNSGQSAIKHNVKSLITG
 PSKLSRGMPLPOTLEIPENIKVVERKVEDYKAGETVRSRHTSVVSSGSPVLSRSTLHEAP
 KAQLSGIIDDTSARTPVSYQNTMSRGSPPMNRSTSDVTISNNKSNHKKSLTPTQ
 RESIPAKSPVGVDPVVSHPSPFDPHHRGTAGEVRSHPDLTPDAMPFHRLADPAA
 AYLFQRLSDPTFGPSQOYLAMENTROTILANDYITSQOMVNLRDPDVARGLSPRQP
 LGLYPATRGIIDLTNMPPTILVHPGCTSTPMDRITVPGTQIFPRPYPNYSAGMS
 PGHPHTLAAASAEIERERERERERERERERERERERERERERERERERERERER
 SPSTVQETMLQRPVSFQGTNGTSVITPLDPTAQLRIMPLPAGGSPISQGLPASRYN
 TAADALVAADAASAPQMDVSKTSEISRRHYETPSDALEVISPASSAPPEKQUTY
 QPEVKAANQADPTROYEGPLHYPQOESPSPQOOLPPSQOAGMQGVPRTHLIT
 LADHICOIITQDFARNQVSQTPQPTSTFQNSPALSVPVTRTNSRNPVSESOAQ
 SVHQRPGSRSPENLVDSRGRSGKSPERSHVSSEPEVPIPPPOVPVHVKEKODSL
 LLSORCAEPAEORNDARSIGSYLPSFTKLENTSPMVKSKQOEIFRKLNSGGGDS
 DMAAQPCTEINLNPATVTSVSSSRGSPADPNLGHLEDIRKALMSFDPKVEDH
 GVMSVPMGVPGTANTSVVTSGERHSDPSPHSGVCKPKLISKMSRKSRSKSP
 GQGLYTERPSSVSHSEGDYHROPQWAWEDRPSSTGSTQFPFNPMLTMRSLSTPP
 TPIACAPSVAQNQAAPLQONRIWEREPAPLLSAQYETLSDSD"

ORIGIN

Alignment Scores:
 Pred. No.: 5,63e-77 Length: 7949
 Score: 4194.00 Matches: 1097
 Percent Similarity: 53.92% Conservative: 340
 Best Local Similarity: 41.16% Mismatches: 720
 Query Match: 31.74% Indels: 508
 DB: 9 Gaps: 108

US-09-522-753-5 (1-2517) x AB028970 (1-7949)

123 LeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThr 142
 447 -----GAAGGGCTGAGGGCT---TCTGCAGATGCTAAGAGGATCCAGCATTCGGA 494
 143 GlyLysLeuGlu---ProValSerProProProProHisThrAspProGluLeuGlu 161
 495 GGCACAAACATGAAGCTCCATCTCTCCAAATTCGGGGCAACCATGTGGAGATGATCAAAAT 554
 162 LeuValProProArgLeuSerLysGluLeuLeuLeuGlnAsnMetAspArgValAspArg 181
 555 GCTTACCTTCAAACTCTCAAGAGGAGAGTAAATACAGAGTATGGATCGTGTAGATCGA 614
 182 GluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeuGlu 201
 615 GAAATTCACAAAGTAGAACAGCAGATCTCTTAAACTGAAAAGAAACAAACACAGCTTGA 674
 202 GluGluAlaAlaLysProGluProGluProGluLysProValSerProProProLysLeuSer 221
 675 GAAGAGCAGCTAAACCTCTGAGCGTGAAGAGCCGCTGCTCCCTCTCTGTGGAGCAG 734
 222 LysHisArgSerLeuValGlnIleLeuThrAspGluAsnArgLysLysAlaGluAlaAla 241
 735 AACACCCGAGTATTGTCCAAATTTATTAATGATGAGATCGGAAGAAACAGAGAAGCT 794
 242 HisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuThrAsnGlnProSer 261
 795 CATAAAATTTTGAAGGTCTTGCCCAAAAGTTGAATCCCATCTGTATAACAGCCATCA 854
 262 AspThrArgGlnThrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIle 281
 855 GATACCAAGCTGTACCATGAGAACATCAAGACAAACAGGTGATGAGGAAAAAATCACT 914
 282 LeuThrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArg 301
 915 TTATTTTTTAAAGAAAGAAATCATGCAAGAAACAAAGGAGCAAAAAATCTGCCAGCGT 974
 302 TyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArg 321
 975 TATGATCAGCTCATGGAGGCGATGGGAGAAAAAAGTGGACAGAAATAGAAAAATCTCGG 1034
 322 ArgArgAlaLysGluSerLysValArgGluThrTyrGluLysGlnPheProGluLysLeuArg 341
 1035 AGGAAAGCTTAAAGAAAGCAAAACAAAGGGAATATCTATGAAAGCAGTTTCCAGAAATTCGA 1094
 342 LysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyClnArgGlySerGlyLeu 361
 1095 AAACAAGAGAACAGCAAGAAAGATTTCAG---CGAGTTGGGCGAGGAGGAGCTGTCTT 1151
 362 SerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGlu 381
 1152 TCAGGCCACCATTTCTAGGAGTGCAGATGAGATTTCTGAAATTTATTGATGGGCTCTCTGAG 1211
 382 GlnGluAsnLeuLysGlnMetArgGlnLeuAlaValIleProProMetLeuThrAsp 401
 1212 CAGGAGAAATTAATGAGAAACAAATTCGGCAGCTCTCTGTGATTCACCTATGATTTTGTAT 1271
 402 AlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLys 421
 1272 GCAGAACAAAGACAGTCAAGTTTCAATTAATGATGATGGCTTATGGAGACCTATGAAA 1331
 422 ValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGlnLysGluThrPheArg 441
 1332 GTGTATAAGATAGGCAGTTTATGAATGTTTGGACTGCACATGAAAGGAGATCTTTAAG 1391
 442 GluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLys 461
 1392 GACAAAGTTTATCCAGCATCCAAAAAATCTTGGCAATTAATTCATCATTTGGAGAGGAG 1451
 462 ThrValAlaGluCysValLeuThrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSer 481
 1452 AGTGTCTCGATTGTGTTTGTATTACTATTATTAACCAAGAAAAAATGAGAAATATAAGCC 1511
 482 LeuValArgArgSerTyr---ArgArgArgGlyLysSerGlnGlnGlnGlnGlnGln 500

Db	5325	-----		-----CCTGCCACCTGGCAGT	5342
Qy	1854	HisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAla---	LeuGln	1872	
Db	5343	CATGGATATGTTCCTCCCTCCCT---TCAGTAAGAACTCAGGAGCAACTGTGGCAA	5399		
Qy	1873	GlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluPro	1892		
Db	5400	CAGAGACCAGTGTTTCCAAGGAACAATGGAACCATGTGAATCACACCTTTGGATCCA	5459		
Qy	1993	SerLysProThrValLeuArgSerThrSerThrSerProValArgProAlaAlaThr	1912		
Db	5460	ACT---GCTCAGCTACGA-	5483		
Qy	1913	PheProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyProThrLeu	1932		
Db	5484	-----CTGCCTGTGGGGC-	5498		
Qy	1933	MetGluProValLeuLeuProLysGluAlaProArgValalaArgProGluArgProArg	1952		
Db	5498	-----	5498		
Qy	1953	AlaAspThrGlyHisAlaPheLeuAlaIlyeProProAlaargSerGlyLeuGluProAla	1972		
Db	5499	-----CCTTCAATAAGCCAACGCCCTG-	5519		
Qy	1973	SerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAla	1992		
Db	5520	-----CCAGCCTCCCCTTACAAC	5537		
Qy	1993	ThrIleAlaArgThrProAlaLysAsnLeuAlaProHisAlaSerProAspProPro	2012		
Db	5538	ACTGCTGGATGCGCTGCTCTTGTG-	5567		
Qy	2013	AlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSer	2032		
Db	5568	-----GATGCTGCAGCTTCTGCACCCAGATGATGTGTCCAAAACAAAGAAGATATCT	5621		
Qy	2033	IleGlnGluLeuGluLeuArgSerLeuGlyTyHisGlySerTySerProGluGly	2052		
Db	5622	TCTCACAGGTATGAACACCTAGC-	5651		
Qy	2053	ValGluProValSerProValSerSerSerLeuThrHisAspLysGlyLeuProLys	2072		
Db	5652	ATTGAGTGATAAGTCTCTGCAGCTACCTGGCCACCACCCAGGANAACATGCAGACCTAT	5711		
Qy	2073	HisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGly	2092		
Db	5712	CAGCCAGAGGTTGTTAAGGCAATCAAGCGCAAAATGATCCTACCAGACAATATGAAGGA	5771		
Qy	2093	ProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSer	2112		
Db	5772	CCA-----TTACATCACTATCGACCA-	5795		
Qy	2113	GlnProSerSerProLeuLeuGlnThrAlaProGly-	2125		
Db	5796	CAGAAATCACCATCTCCCAACACACAGCTGCCCTCTTCACAGGCAGAGGAATGGGG	5855		
Qy	2126	---ValLysGlyHisGlnArgValValThrIleAlaGlnHisIleSerGluValIleThr	2144		
Db	5856	CAAGTGCACAGGACCCATCGCTGATCATCACTTGCTGATCATCTCTGTCAAAATATCACA	5915		
Qy	2145	GlnAspTyThrArgHisHis-	2159		
Db	5916	CAAGATTTTGTAGAATAATCAAGTTTCTCGCAGACTCCCCAGAG-	5966		
Qy	2160	ProAlaProLeuTySerPheProGlyAla-----SerCysProValLeuAspLeuArg	2177		
Db	5967	ACTTCTACATTCAGAACTACCTTCTGCTTGGTATCTACACTGTG-	6017		
Qy	2178	ArgProProSerAspLeuTyLeuProProProAsp-----HisGlyAla	2192		

Db	6018	ACTAAACATCAAAACCGTTACAGCCCAAGAAATCCAGGCTCAGTCTGTGTCATCATCAAGA	6077
Qy	2193	Pro---AlaArgGlySerProHis-----SerGluGlyGlyLysArgSer	2206
Db	6078	CCAGGTTCAAGGCTCTCTCCAGAAAATCTGTGCACAAATCCAGGGGAAGTAGGCGCTTGA	6137
Qy	2207	ProGluProAsnLysThrSerValLeuGlyGlyGluAspGlyIleGluProValSer	2226
Db	6138	AAATCCCGACAGAGAGAGTCAAGTCT-----TCTTCGGAGCGCTACGAGCCCATCTCC	6188
Qy	2227	ProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyr	2246
Db	6189	CCACCCAG-----GTTCCGGTGTGCATGAGAAACAGGACAGCTGTGCTCTTCTCT	6242
Qy	2247	ArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSer	2266
Db	6243	CAGAGGGCGCAGAGCTGCAGACAGAGAAATGATGCCCGCTCACCAGGGAGTATAGC	6302
Qy	2267	GlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLys	2286
Db	6303	TACTTGCTTCACTTCTTCAACAGCTT---GAAATACATCACCCATGGTTAAATCAAAG	6359
Qy	2287	LysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIle	2306
Db	6360	AAGCAGAGAGATTTTCGTAAATGTAATCTCTCTGGTGAGGTGACTCTGATGGCAGCT	6419
Qy	2307	SerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThr	2326
Db	6420	GCTCAGCAGGAACTGAGATCTTTAATCTGCCAGCAGTTACTAGCTCAGGCTCAGTTAGC	6479
Qy	2327	TyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIle	2346
Db	6480	TCTAGAGGCCATCTCTTTGCTGATCTCTGCCAGT---AATCTGGCTGGAAGACATTATC	6536
Qy	2347	ArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGlu-----Ser	2360
Db	6537	AGGAAGCTCTCATGGGAAGCTTTGATGACAAAGTTGAGGATCATGGAGTTGTCATGTCC	6596
Qy	2361	ProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAla	2380
Db	6597	CAGCTATGGAGTA-----GTGCTGTGTTACTGCCAACACCTCAGTT-----	6638
Qy	2381	MetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGly	2400
Db	6639	-----GTGACCAGTGTGTGAGACACGAAGAGAGAGAGGGGACCCATCACCTCATTCAGGA	6692
Qy	2401	Gly-----LysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAla	2418
Db	6693	GGAGTTTGCAAAACCAAGCTGTATCAGCAAGCTCAACACGACGAGAAATCTAAGTCTCTATA	6752
Qy	2419	Pro-----GlyLeuAlaSerGlyAspArgProProSerValSerValHisSerGlu	2436
Db	6753	CCTGGCAAGGCTACTTAGGAACGGAACGGCCCTCTTCAGTCTCTCTGTACATTTCAGAA	6812
Qy	2437	GlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSer	2456
Db	6813	GGGATTAACCATAGGCAGACGCCA-----GGGTGGGCTGGGAGACAGCGCCCTCTTCA	6866
Qy	2457	AlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMet	2476
Db	6867	ACAGGCTCAACTCAGTTTCTCTTATAACCCCTCTGACTATGCGGATG-----CTCAGC	6917
Qy	2477	AlaSerProProProGlyLeuProAlaGlySerGly---ProLeuAlaGlyProHis	2495
Db	6918	AGTACTCCACACACCATGGATTGATGTGCTCCCTCTGCGGTGGAACCAAGCAGCTCTCAC	6977
Qy	2496	-----HisAlaTrpAspGluGluProLysPheProLeuLeuCysSerGlnTyrGluThr	2512
Db	6978	CAACAGAACAGGATCTGGGACGAGAGCCTGCCCACTGCTCTCAGCACAGTACGAGACC	7037
Qy	2513	LeuSerAspSerGlu	2517
Db	7038	CTGTGCGATAGTAT	7052

RESULT 9	MMU35312	7780 bp	mRNA	linear	ROD 19-OCT-1996
LOCUS	MMU35312				
DEFINITION	Mus musculus nuclear receptor co-repressor mRNA, complete cds.				
ACCESSION	U35312				
VERSION	U35312.1				
KEYWORDS	GI:1022717				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	1 (bases 1 to 7780)				
AUTHORS	Horlein,A.J., Naar,A.M., Heinzel,T., Torchia,J., Gloss,B., Kurokawa,R., Ryan,A., Kamei,Y., Soderstrom,M., Glass,C.K. and Rosenfeld,M.G.				
TITLE	Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor				
JOURNAL	Nature 377 (6548), 397-404 (1995)				
MEDLINE	96008539				
PUBMED	7566114				
REFERENCE	2 (bases 1 to 7780)				
AUTHORS	Horlein,A.J., Neer,A.M., Heinzel,T., Torchia,J., Gloss,B., Kurokawa,R., Kamei,Y., Ryan,A., Glass,C.K. and Rosenfeld,M.G.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-SEP-1995) Andreas J. Horlein, School of Medicine, UCSD, 9500 Gilman Drive, La Jolla, CA 92093-0648, USA				
FEATURES	Location/Qualifiers				
source	1..7780				
	/organism="Mus musculus"				
	/mol_type="mRNA"				
	/db_xref="taxon:10090"				
	/tissue_type="pituitary"				
	117..7478				
	/notes="N-CoR"				
	/codon_start=1				
	/product="nuclear receptor co-repressor"				
	/protein_id="AAB17125.1"				
	/db_xref="GI:1022718"				
	/translation="MSSSGYPPNQAFSTEQSRYPHSVQYTFPSRHQOEPAVDPYR				
	SSHLEVSQQLQDQOQQLRRRLSLEFHPGSDRQPRRRSGVGEFHPGSPVDHD				
	SLEKPRLEQVDSDFHFIISAAFLVFLHPEGLSSANAKKDPFVYKHAEPVSP				
	SGQCPGDQVSPKSLSEKELIYDMDRVDREIAKVEQILKLKKQQLQEEAAKPE				
	PEKPPVPPVEQKHSIIVIIYDMDRVDREIAKVEQILKLKKQQLQEEAAKPE				
	ENIKTNVKKLILFKRNHARKOEKI CORVDOLMEANEKKVDRIENNPRKAK				
	ESTRIYKQFPEIKRQEQRFQVQGRGLSATIARSEHSEIILDLGSEQEN				
	NEKQKQLSVIPPMFDASQRRVKFINMGLMEDPMKVKDQFMNVWTDHKEIFPK				
	KFTQHPNGLIASYLERSVPDVLVLYLTIKNENYKALVRNRYKRRGRNQIARP				
	SOBEKVEEKEDEKAEKTEKEDKEDKEDSKETTKEDKRTAETAEPEPEEQ				
	VTPGRKTANSQGRGKRVTRSNMTSEAAAANAATAEPPLPPPPPISTEPVET				
	SWTEEMEVAKGLVHGRRNNAIAKMGVTKSEACKNFYENYKRNLDLLOQH				
	QKASRPREDVDQCESVASTVSAQEDEDI EASNEENPEDESGAENSDDTESAP				
	SPVEAKSSEDSSENAASRGNTFPVALEATDTPAPCASPPSAVPTTKPAERSEV				
	VTDSAAETAPMDVDHEECAGSVLDPFAPTKADSDVDPENQVPENTASKEGDAK				
	ERLESTSEKTEARDVDVAEQIERPEPQSDSDSATCSADEGVDEPERQVFPMD				
	AKPLSLTPGSLILISSP IKPNLLDLPOLQHRAAVIPMVSCCTPCNIPITGPVGVALY				
	QBHIKMHESALLLEQROQEOVDLECRSTSPCSTKSPNREWEVLQAPHQVITNL				
	PEGVRLPTRTPRPPPLPSSKTIVASEKPSFIMGSSISQGTGGYLSHNAQVQE				
	APKPSVGSISLGLPQKSTKAPLTYIKOEFSPPQSQNSQEGGLVRAQHEGVVGT				
	AGAVQSGSTIRTPGASKIVETISLSRGSITQTPALPQAGITPAELVKGPSVMP				
	ESSEPKVREAAKASIKVYEGKSHILSYDNIKNAREGTRSPRTAHMSLKRSEAVE				
	GSTIKQCMWRESVPSAPLEGLICRALPRGSPHSDLKERTVLSGIMQGTPTATASEF				
	DGLYKPKQIKRSPPIRFAEGATKPKPYDGLTIKEMGRSIEHPRODILTOESRFT				
	PEVQSTRILIEGSIQGPPIKFDNNSSGQSAIKHNVKSLITGPKSLPRGMLSEIVENI				
	KVVERKGVEDVAPVPRARHTSVSSGSPSVLSTLHEAPKQSLPGLYDSSSARKTP				
	VSQNTISRGSPMMNRSTSVSSKSHERKSTLTPTQRESIPAKSPGVDPVIVSHS				
	PDFPHRRSAGVYRSHLPDTHLDPAFTRGLDPAAYLLQRLQSLPTPCYPSQVLY				
	AMENTROTLLNDYITSQOMVNLRPDPVTRGLDPREQPLGLYPATRGITDLTNMPTPI				
	LVPFHAGSTPMDRLTYIPCTQVTPPPRYNAASISPCGHPHLLAAASAEREER				
	EKERERERERERERERERERERERERERERERERERERERERERERERERER				
	QKPSVFGTNGTSVITPDTPAQRLMELPFGGPPSISQGLPASRINTAADALALVD				
	AAASAPMDVSHASKESKHAARLEENLRSSAAVSEQQLEQKNEKRSVQCVCVS				
	SALPSGKAQPHASVVYSEAGKQKGGPPPKRSRYEBELTRGKTTITANFNITDVIITQIA				

SDKADERGSQSSSSSSSHRYETASDAIEVISPASSPAPPEQKPOAQYQDMVKAN					
OAENESTROYEGLEPHYRSQSPSPQOPLPSPSSQSGMGOVPRTHLITLADHIC					
OIITODEARNOVPSOASTSTFOTSPALSSTPVRTKTSRYSVESQSOVLHPRGPR					
VSPENLVKSRGSRKPKPERSHIFSEPEPIPSPEQRAVHKQSKMLLSQSGVDPA					
EQRSDSRPGSISYLPSPSTFKLESTSPMWKSKQIIFRKLNSGGSDSDMAAAQPGTE					
IFNLPVATTSAGVSRFSFADPASNLGLEDIIRKALMGSDFKVEDHGVMAWHPYGI					
MPGSASTSVTTSSEARDEGEPHAGVCKPKLINSKRSKSPITPGQSYLGTERTPS					
SVSSVHSGDVHROTPTGMAWEDRPSSTGTSTQFPYNPLTIRMLSSPTPTQIACAPSAIT					
QAAHQONRIWEREPAPILLSAQYETLSDSD					

Alignment Scores:	7.66e-77	Length:	7780
Pred. No.:	4187.00	Matches:	1104
Score:	53.94%	Conservative:	347
Percent Similarity:	53.94%	Mismatches:	798
Best Local Similarity:	41.04%	Indels:	442
Query Match:	31.68%	Gaps:	104
DB:	10		

US-09-522-753-5 (1-2517) x MMU35312 (1-7780)

QY	16	GluProArgTyrProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThr	35
Db	165	CAGAGTCGCTACCCCTTCGCACCTCGTTTCAGTACACCTTTCCAGCGCCCTCACCAGCAG	224
QY	36	AspValGlyLeuLeuGluTyrGln-----HisHisSerArgAspTyrAlaSerHisLeu	53
Db	225	GAATTGGAGTCCTGACTACCGTCTTCTCATCTTGAAGTCAGTCAGCGGTCACAGCTC	284
QY	54	SerProGlySerIleIleGlnProGlnArgArgProSerLeuLeuSerGluPheGln	73
Db	285	TTGCAGCAGCAGCAGCAGCAGCAGCTTCGACAGCGGCTTCCTGCTTTCCGAGTTTCAC	344
QY	74	ProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeuPro	93
Db	345	CCGGGTTCGACAGGCCCCAGGAA-----AGGAGAAGTGGATACAGCAGTTCCACCCG	398
QY	94	GluLeuGlyLysSerGlnMetGluPheIleGluSerLysArgProArgLeuGluLeu	113
Db	399	GGCCCTTCCCGGTGGACCATGACTCGCTGGAGTCCAGCGGCTCGCTGGAGCAGGT	458
QY	114	-ProAspProLeu-----LeuArgProSerProLeuLeuAlaThrGlyGlnPr	129
Db	459	TCCGACTCCCACTTCCAGCGCATCAGTGTGCTGCTCTCCCTTTGGTGACACG-CTGCC	517
QY	129	alaGly-----SerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGl	146
Db	518	AGAAGGACTGAGGTCTTCTGCCAATGCTAAGAAGGATCCGGCATTTGGAGTCAACATGA	577
QY	146	u---ProValSerProProHisThrAspProGluLeuGluLeuValProPr	165
Db	578	AGTCTCTTCTCTCCCTCTCTGGGAGCATGCGGAGATGATCAGATGCTCCTCCTTC	637
QY	165	oArgLeuSerLysGluGluLeuIleAsnMetAspArgValAspArgGluIleThrMe	185
Db	638	AAAACTGTCAAAGGAAGAGCTGATACAGAGCATGATCGTGTGTCGAGAGAAATTCGAA	697
QY	185	tValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeuGluGluGluAla	205
Db	698	AGTAGAACAGCAGATCTCTTAAACTGAAAAAAGAGCAGCAACAGCTCGAAGAGAACTGC	757
QY	205	alysProProGluProGluLysProValSerProProIleGluSerLysHisArgSe	225
Db	758	TAAACCCAGAGCCCTGAGAACCTGTGTCCTCTCTCCGTCGAGCAGCAGCAGCAG	817
QY	225	rLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIle	245
Db	818	TATTGTCCAAATCATTTATGACAGAAATCGGAAAAAAGCAGAGAGCTCATAAATAAT	877
QY	245	uGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGl	265
Db	878	TGAAGGTCTTGGCCCAAAAGTTGAACCTGCCCTCTACTCAACACCGCCCTCAGATACCAAGGT	937

QY 265 nTyrHisGluAsnIleLysIleIleGlnAlaMetArgLysLysIleLeuTyrPheLy 285
DB 938 GTACCCAGAGAAATCATCAAGCAACACCGAGTGGATGAGGAAAAAACTCATTTTATTATTAA 997
QY 285 sArgArgAsnHisAlaArgLysGlnTTrpLysGlnLysPheCysGlnArgTyrAspGlnLe 305
DB 998 AAGAGAAATCATGCAAGAAACAAAGGAAACAAAATAATCTGCCAGCTTATGATCAGCT 1057
QY 305 uMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgAlaLy 325
DB 1058 CATGGAAGCATGGAGAAAAAGTGGACAGAAATAGAAAAATAATCTCGGAGGAAAGCTAA 1117
QY 325 sGluSerLysValArgGluTyrTyrGluLysGlnPheProGluLleArgLysGlnArgG1 345
DB 1118 AGAAAGCAAAACAAAGGAAATATATGAAAAACGAGTTTCCAGAAATTCGAAACAAAGAGA 1177
QY 345 uLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAl 365
DB 1178 ACAGCAAGAAAGATTTCAG---CGAGTTGGTCAAGGGAGCTGGTCTTTACGCCACCAT 1234
QY 365 aAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGlnAsnLe 385
DB 1235 TGCTAGGAGTGAGCATGAGATTCTCGAAATATTATGATGGTCTTTCGAACAGGAGATAA 1294
QY 385 uGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnG1 405
DB 1295 TGAGAAAGCAAAATCGTCAGCTTTCTGTGATTCACCTATGATGTTGATGATGATGATGAT 1354
QY 405 nArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAs 425
DB 1355 AAGGGTCAAAATTCATCAATATGATGAGTGGCTGATGGAGATCCCAATGAAGTTTATAAAGA 1414
QY 425 pArgGlnValMetAsnMetTrpSerGluGlnGlnLysGluThrPheArgGluLysPheMe 445
DB 1415 CAGACAGTTTATGATGTTTGGACTGACCATGAAGAGGAGATCTTTAAGGACAGTTTAT 1474
QY 445 tGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaG1 465
DB 1475 CCAGCATCCAAAAAACTTTGGACTTAATTGCACTCTATTTGCAAGGAAAGAGTGTCTCTGA 1534
QY 465 uCysValLeuTyrTyrTrpLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgAr 485
DB 1535 TTGTGTTTATATTACTATTATTAACCAAGAAATAATGAGAAATTAAGAGCCCTCTGTGAAG 1594
QY 485 gSerTyr---ArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnG1 504
DB 1595 GAATTATGGAACAGCAGAGCAGAAATCAGCAGATTGCCCGTCCCTCCACAAAGAAAA 1654
QY 504 nGlnGlnGlnGlnGlnProMetProArgSerSerGlnGlnGlnGlnGlnGlnGlnG1 524
DB 1655 AGTAGAAGAAAAAG-----GAGAGGATAAAGCAGAAAAAAC 1690
QY 524 uLysGluLysGluAlaGluLys-----GluGluGluLysProGluValGluAsnAspLy 542
DB 1691 AGAGAAAAAGGAGAAAGAAAGGATGATGAAGAAAAAGATGATGAAGGAAAGACTCTAA 1750
QY 542 sGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAsnAspGluLysG1 562
DB 1751 AGAAACAAACAGAAAAAGCAGACGGAAGCCACAGCAGAGAACCTGGAAGAGAGA 1810
QY 562 uAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgI1 582
DB 1811 GCAGGTCACTCCAAGGGGGGAAAGACTGTCTAACAGCCAAAGGCCGCGGGAAGGCCGGGT 1870
QY 582 eThrArgSerMetAlaAsnGlu-----AlaAsnSerGluGluAlaIleThrProG1 599
DB 1871 CACCAAGTTCGATCAACAGTGAAGCTGCAGCTGCCAATGCTGCTGCTGCAAGCCACTGAGGA 1930
QY 599 nGlnSerAlaGluLeu-----AlaSerMetGluLeuAsnGluSerSe 613
DB 1931 GCCCCCGCCACCCCTCGCCGCCACACAGAGCCCATTTCTACAGAACCTGTTGAGACTTC 1990
QY 613 rArgTrpThrGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAs 633

DB 1991 AAGATGGACAGAAAGAAATGGAAGTTGCTAAAAAAGGCGCTGTGAGAACATGTCGTAA 2050
QY 633 nTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheTy 653
DB 2051 CTGGCAGCATTTGCTAAATGTTGGAACTAAAGTGAAGCCAGCTGTCAAAAACTTCTA 2110
QY 653 rPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMe 673
DB 2111 TTTTAACTATAAAAGACGGCATTAATCTTGACAACTTTTGCAGCAACATAAACAAGAAC 2170
QY 673 tGluLysGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaAlaSerGluG1 693
DB 2171 TTCACGGAAACCCCGTGAGGAACGAGATGATCTCAGTGTGAAGTGTTCCTCCACTGT 2230
QY 693 uAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGlyValSerGlyAs 713
DB 2231 TTCCTGCC-----CAAGAGATGAGATATTGAAGCTCA-----AA 2266
QY 713 nGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProAr 733
DB 2267 TGAGGAGGAAAAATCCAGAGATAGTGAAGT----- 2297
QY 733 sGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSe 753
DB 2298 -----GCTGAAAAATAGTTCTGTATACAGAAAAGTGTCTCCCTC 2332
QY 753 rProHis---ThrGluAlaAlaLys-----AspThrGlyGlnAsnGlyProLysPr 769
DB 2333 TCCTTCACAGTTGAAGCTGCCAAGTCCAGTGAAGACAGCAGCGTGAATAATGCT----- 2384
QY 769 oProAlaThrLeuGlyAlaAspGlyProProGlyProProGlyProProArgArgTh 789
DB 2385 ----- 2386
QY 789 rSerArgAlaProIleGluProProAlaSerGluAlaThrGlyAlaProThrProPr 809
DB 2387 TTCTCGAGAAACACCCAGCCTGTGGCTTGAGCTTGAGGCCACCCTGACCCCTGCACCTG 2446
QY 809 oProAlaProProSerProSerAlaProProValValProLysGluLysGluG1 829
DB 2447 TGATCTCC-----TCCTTCAGCTTCCACCAACCAACAGCAAGAAAGGAAAGCGT 2500
QY 829 uGlu-----ThrAlaAlaAlaProProValGluGluGlyG1 841
DB 2501 GGAGGCCAGGTGACCGCAGCGCCAGTGCAGAGACCGCAGACCGATGCGACGTAGACCA 2560
QY 841 uGluGln-----LysProProAlaAlaGluGluLeuAl 852
DB 2561 TGAGGAGTGCCTGCGCGAGGCGAGTCTGTCTTGTATCCACAGCCCTTACCACAA----- 2615
QY 852 aValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluG1 872
DB 2616 -GCCGACTCCGTGGACCCAGAAATGAGGTA---CCAGAAATACTGCGTCTAAAGGTGA 2671
QY 872 uGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLy 892
DB 2672 AGGGGATGCCAAGAAAGAGACTTGGAGAGAC----- 2705
QY 892 sAlaGluLysLysGluGlyCysArgAlaThrThrAlaLysSerSerGlyAlaPr 912
DB 2706 -AGTGAGAGACAGAGCGCTAGAGATGAAGACGTGGTAGTGGCTGAACACAGATCGAGAGGCC 2764
QY 912 oGln-----AspSerAspSerAlaThrCysSerAlaAspGluValAspGluAl 929
DB 2765 TGAGCCACAGTCCAGACGACTCCAGTGCCTTCAGTGCAGTGCAGATGAGGGTGTGGATGG 2824
QY 929 aGluGlyGlyAspLysAsnArgLeu-----LeuSerProArgProSerLeuLeuThrPr 947
DB 2825 AGAG---CCAGAGAGCAGAGAGTGTTCCTCATGGATGCAAGACCTTCATTGTTAACTCC 2881
QY 947 oThrGlyAspProArgAlaAsnAlaSerProGlnLysPro-----LeuAspLeuLysG1 965

Qy	1621	laspLeuTyArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyI	1641
Db	4808	AGAGGTTTATCGAGCCACCTTACCACCGCACTTGGATCCA---GCTATGGCTTTTCACAG	4864
Qy	1641	eProLeuAspAlaalaalaIaTyTyLeuProArgHisLeuAlaProAsnProThrTy	1661
Db	4865	GGCTTTGGATCCTGCTGCTTACCTTACAGAGACAGCTTTTCCAAACCCAGGATA	4924
Qy	1661	rProHisLeuTyTyProTyTyLeuIleArgGlyTyProAspThrAlaalaLeuGluAs	1681
Db	4925	CCCAAGTCAGTACCAGCTCTAT-----GCATGGAGAA	4957
Qy	1681	n---ArgGlnThrIleIleAsnAspTyTyIleThrSerGlnGlnMethHisHisAsnThrAl	1700
Db	4958	TACAAGGCAGACAATCTCTCAACGATATACATTACCTCCACAGCAGATGCGGTGAATCTG-	5015
Qy	1700	aThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLe	1720
Db	5016	-----CGCCCTGTGATGCACAGGGGACTGTGTCCTCCACGAGAGCAGCCACT	5059
Qy	1720	uAlaLeuAsnTyTyAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLe	1740
Db	5060	GGGCTCCCTTACCAGCT---ACAAGAGGAATCATTTGACCTGACCAATATGCTCCA--	5114
Qy	1740	uProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTy	1760
Db	5115	-ACAATCTTAGTGCTCATGCAGGGGAAACGAGCACCCCTCCCATGCACAGGATCACGTA	5173
Qy	1760	rLeuProThrAlaProGlnProPheSerSerArg--HisSerSerSerProLeuSerPr	1779
Db	5174	TATTCCCTGGTTACACAGGTTACTTCCCTCCAGGCCATATAACGCTGCTTCTGTGCTCC	5233
Qy	1779	oGlyGlyProThrHisLeuThrLysProThrThrSerSerGluArgGluArgAs	1799
Db	5234	AGGACCCCAACACACTT-----GCAGCAGCTGCAAGTGTCTGAGGGAGACGAGA	5284
Qy	1799	pArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrTh	1819
Db	5285	ACGGGAAGGGAGAGAGCGGACGCGAACTGTCAGCGCGAGCTGAGCGCTGAACCTGAAC	5344
Qy	1819	rThrValGluHisAlaPro-----IleTrpArgProGlyThrGluGlnSerSerGl	1836
Db	5345	AAGGATCGCTGCTGCTCCGCTGACCTCTACCTACGACCAAGTTTCAGAACAG-----	5396
Qy	1836	ySerSerGlySerSerGlyGlyGlyGlyGlySerSerArgProAlaSerHisSerHi	1856
Db	5397	-----CCAGGCGCTCGCAGCACCGGATA	5422
Qy	1856	sAlaHisGlnHisSerProIleSerProArgThrGlnAspAla---LeuGlnGlnArgPr	1875
Db	5423	TGTTTCGCTCCCTTCCCT---TCAGTAAGAACTCAGGAGACCATCTCTGCAACAGAGACC	5479
Qy	1875	oSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysPr	1895
Db	5480	CAGTGTTTTCCAGGCGCAACCAATGAAACGAGTGTAATCACCTTTGAGCCCACTGTCTCA	5539
Qy	1895	oThrValLeuLeuArgSerThrSerThrSerSerProVal-----ArgProAlaAl	1911
Db	5540	GCTACGATCATGCCACTGCTCTGGGGCCCTTCCATAAGTCAAGCGCTGCCAGCTC	5599
Qy	1911	aThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyProTh	1931
Db	5600	CCGTTTACAACACTGCTCGGATGCC---CTGGCTGCTCTTGTGATGCTGCAGCTTCTGC	5656
Qy	1931	rLeuMetGluProValLeuLeuProLysGluAlaProArg-----	1944
Db	5657	ACCCAGATGGATGTTTCCAAAAACAAAGAGAGTAAGCATGAAGCTGCCAGGTTAGAAGA	5716
Qy	1945	-----ValAlaArgProGluArgProArgAlaAspThrGlyHisAl	1958
Db	5717	AAATTTCAGAGCAGGTCAGACAGCATGTAGTAACACAGCAGCAGCTAGACAGAGAAAACCT	5776
Qy	1958	aPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGl	1978

Db	5777	GGAGGTGGGAAGAGAGATCTGTTCAAGTGTGTGTGCACTTCTTTCAGCCCTTCCAAAGTGCGCAA	5836
Qy	1978	ySerGluProArgProLeuVal	1987
Db	5837	GGCCCGCCCTCAATGCTCAGTAGTGATTCCTGAGGCTGGGAAGATAAAGGGGCTCTCTCC	5896
Qy	1987	oValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsn	2005
Db	5897	AAAATCCAGATATGAGGAAGAGCTAAGGACCCGAGGGAAGACTACCATTAATCGCAGCTAA	5956
Qy	2005	sHisAlaSerProAspProAlaProProAlaSerAlaSerAspProHisArgGluLys	2025
Db	5957	CTTCATAGAGCTGATCATCAACCCGCGAAATGCTCTCGACAAAGGATCGGAGGAACCGTGG	6016
Qy	2025	sThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyHisGln	2045
Db	6017	CTCTCAAGTTCAGACTCTTCT	6058
Qy	2045	ySerSerTySerProGluGlyValGluProValSerProValSerSerProSerLeuTh	2065
Db	6059	GTATGAACCGCTAGTGCATCCATTGAGGTGATAAGTCCCGCCAGCTCACCTGCACCACC	6118
Qy	2065	rHisAspGlyGlyLeuProLysHisLeuGluLeuAspLysSerHisLeuGluGlyGln	2085
Db	6119	CCAGGAAAGCCACGAGCCCTATCAGCCAGACATGGTTAAGGCAATCAACGACGAAATGA	6178
Qy	2085	uLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaHisLeuProHis	2105
Db	6179	GTCACATCGACAGTATGAAGTCCA	6211
Qy	2105	sLeuArg	2120
Db	6212	TTATCGGTCCCGACGAGAATCACCATCTCCACAGCAACAGCCACCATGCGCCCATCTTC	6271
Qy	2120	uGlnThr	2138
Db	6272	CCAGTCAGAGGGAATGGGACAGGTGCCCGACGATCCCATCGACTGATCACACTTGTGTACCA	6331
Qy	2138	sIleSerGluValIleThrGlnAspTyrThrArgHisHis	2157
Db	6332	CATCTGCAATATATACACAAGATTTTGTGAATAATCAAGTTCCTTCGCGAGGCTTCTAC	6391
Qy	2157	aProLeu	2173
Db	6392	TTCTACATTCGAAATTCACCATCTGCTTGTTCATCCACACT	6434
Qy	2173	lLeuAspLeuArgProProSerAspLeuTyrLeuProProAsp	2189
Db	6435	GTAAAGAACTAAACCTCAAGCCCTCAGCCCAAGATCAAGTCTCAGACTGT	6487
Qy	2190	HisGlyAlaProAla	2202
Db	6488	CTTGATCCCGACAGGCTCCTAGTCTCTCAGAAATCTTGTTGATAAATCCCGGG	6547
Qy	2202	yGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAspGlyI	2222
Db	6548	AAGCAGCCCTGGAATCTCCAGAGAGAGTCTATC	6598
Qy	2222	eGluProValSerProGluGlyMetThrGluProGlyHisSerArgSerAlaValTy	2242
Db	6599	TGAGCCCATCTCCCAACCCCAAGGCCCTGCTGTG	6652
Qy	2242	rProLeuLeuTyArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerPr	2262
Db	6653	GTTGCTCTTGTACAGAGGGGAGTGGACCTGCTGAGCAAGAGGATGATTCTCGATCACC	6712
Qy	2262	oGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMe	2282
Db	6713	AGGAGTATAAGCTACTTGCCTTCATCTTCACCAAGCTT	6769
Qy	2282	tValLysSerLysGlnGluLeuAsnLysLysLeuAsnThrHisAsnArgAsnGluPr	2302

```

Db 6770 GGTAAATCAAGAAACAGCAAAATTTTCGTAAAGTTGAACCTTCTGTGTGGAGGTGACTC 6829
QY 2302 oGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyTh 2322
Db 6830 TGATATGGAGCTGCTCAGCCAGCAAGATCTTCATCTGCCAGCAGTTACCAATC 6889
QY 2322 rGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLe 2342
Db 6890 AGGTGCAGTGAGCTCAAGAGGCATTTCTTTGTGTGATCCCGCAGT---AACCTTGGTCT 6946
QY 2342 uGluAlaIleArgGlyAlaLeuMetGlyLysTyrAspGlnTTPGluGlu----- 2359
Db 6947 AGAAGACATCATCAGAAAGCTCTCATGGAGAGTTTGTATGATTAAGTTGAAGATCATCG 7006
QY 2360 -----SerProPheLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSe 2376
Db 7007 TGTTCATCTCCATCTCTGTG-----GGCATTTATGCTTGGTGTAGTGCCAGCACCTC 7057
QY 2376 rLeuProAlaAlaMetProIleThrAlaAlaSerGlyArgSerAspHisThrLeuThrSe 2396
Db 7058 AGTG-----GTGACGAGCAGCGAGGCGAGAGATGAAGGGAGGAGCCATC 7102
QY 2396 rProGlyGlyGly---GlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLys 2415
Db 7103 ACCTCATGAGGATGTCACCAAGCTGATCAACAAATCAACAGCAGGAGTCTAA 7162
QY 2415 sSerProAlaProGlyLeuAla-----SerGlyAspArgProProSerValSerSerVa 2433
Db 7163 ATCTCTATTCTCTGGGCAAGCTATTTAGGAACCTGAAGAGCCTTCTCTCTCTCTCTGT 7222
QY 2433 lHisSerGluGlyAspCysAsnArgThrProLeuThrAsnArgValTPGluAspAr 2453
Db 7223 GCATTCAGAGGTGATTTACACAGCGCAGACACCA-----GGATGGGCATGGGAAGATCG 7276
QY 2453 qProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAl 2473
Db 7277 GCCCTCTTCACAGGTTCTACTCAGTTCCTCTTACACCCCTCTGACCATCGGATG----- 7331
QY 2473 aGlyValMetAlaSerProProGlyLeuProAlaGlySerGlyProLeu----- 2491
Db 7332 -----CTCAGCAGTACACCCCTACACAGATCGCATGCGCCCATCTGCCATCACCCA 7384
QY 2492 -AlaGlyProHis-----HisAlaTPAspGluGluProLysProLeuLeuCysSe 2508
Db 7385 AGCAGCTCCATCAACAGAACCCCATCTGGAGAGGAGCCTGCCCGCTCTCTCTCAGC 7444
QY 2508 rGlnTyrGluThrLeuSerAspSerGlu 2517
Db 7445 GCAGTATGAGACATGCTCTGATAGTGAC 7472

RESULT 10
AX578069 7940 bp DNA linear PAT 08-JAN-2003
LOCUS AX578069
DEFINITION Sequence 191 from Patent WO02081745.
ACCESSION AX578069
VERSION AX578069.1 GI:27647277
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1
REFERENCE Garcia, T., roman Roman, S., Baron, R., Call, K., Theilhaber, J.,
AUTHORS Connolly, T., Jackson, A., Bushnell, S. E. and Rawadi, G.
TITLE Genes involved in osteogenesis, and methods of use
JOURNAL Patent: WO 02081745-A 191 17-OCT-2002;
Aventis Pharma S.A. (FR)
FEATURES
Location/Qualifiers
1..7940
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Homo sapiens nuclear receptor co-repressor 1

```

```

(NC0R1), mRNA"
ORIGIN
Alignment Scores:
Pred. No.: 4,9e-76 Length: 7940
Score: 4147.50 Matches: 1099
Percent Similarity: 53.33% Conservative: 341
Best Local Similarity: 40.70% Mismatches: 785
Query Match: 31.38% Indels: 475
DB: 6 Gaps: 108

US-09-522-753-5 (1-2517) x AX578069 (1-7940)
QY 16 GluProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThr 35
Db 289 CAATGCTGTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 348
QY 36 AspValGlyLeuLeuGluTyrGln-----HisHisSerArgAspTyrAlaSerHisLeu 53
Db 349 GAGTTGCGCAGTCCCTGATTATGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 408
QY 54 SerProGlySerIleIleGlnProGlnArgArgProSerLeuLeuSerGluPheGln 73
Db 409 TTGCAGCAACAGCAGCAGCAACAGCTTCGAAGGCGACCTTCTCTCTCTCTCTCTCTCTCT 468
QY 74 ProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeuPro 93
Db 469 CCAGGTTCTGACAGCCCTCAAGAA-----AGGAGAACTAGTTATGAACCGTTTCATCCA 522
QY 94 GluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProArgLeuGluLeu 113
Db 523 GGGCCATCCCGAGTGGATCATGATTCTACTGGAATCGAAGCACCACCGTCTGGAACAGGTT 582
QY 114 ProAsp-----ProLeuLeuArgProSerPro 122
Db 583 TGTGATTCCTATTTCACGCTGTCAGTCTGGCTGCTGGCTTTTGGCTTTAGTGCACCCGTCGCA 642
QY 123 LeuLeuAlaThrGlyGlnProAlaGlySerGluLeuThrLysAspArgSerLeuThr 142
Db 643 -----GAAGGCTCGAGGCT---TCTGAGATGCTAAGAGGATCCAGCATTCGGA 690
QY 143 GlyLysLeuGlu-----ProValSerProProSerProHisThrAspProGluLeuGlu 161
Db 691 GGCACAACTGAAGTCCATCTCTCCATTTTCGGGGCAACCATGTGGAGATGATCAAAAT 750
QY 162 LeuValProProArgLeuSerLysGluLeuLeuIleGlnAsnMetAspArgValAspArg 181
Db 751 GCTTCACCTTCAAAACTCTCAAGAGGAGAGTTAATACAGAGTATGATCGTGTAGATCGA 810
QY 182 GluIleThrMetValGluGlnIleSerLysLeuLysLeuLysGlnGlnGlnGlnGlu 201
Db 811 GAAATTTGCAAAAGTAGAAGCAGCATCTTAAACTGAAAGAAAGAAACAAACAGCTTGAA 870
QY 202 GluGluAlaAlaLysProProGluProGluLysProValSerProProProIleGluSer 221
Db 871 GAAGAGCAGCTTAACCTCTGAGCTGAGAGCCGCTGCTCCCTCTCTCTCTGAGCAG 930
QY 222 LysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysLysLysLysLys 241
Db 931 AAACACCGCAGTATTGCTCAAAATTTATTTATGATGAGATTCGGAAGAAAGCAAGAGCT 990
QY 242 HisArgIleLeuGluGluLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSer 261
Db 991 CATAAATTTTGAAGGCTCTTGCCCAAAAGTTGAACTGCCACTGTATTAACCGCATCA 1050
QY 262 AspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIle 281
Db 1051 GATACCAAGGTGTACCATGAGAACATCAAGACAAACACGAGTGTATGAGGAAAAAATCTCA 1110
QY 282 LeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGlnArg 301
Db 1111 TTATTTTAAAAAGAAATATCATGCAAGAAACAAAGGAAACAAAGGAAATCTGCCAGCGT 1170

```

QY 302 TyrAspGlnLeuMetGluAlaLeuGluLysValGluArgIleGluAsnAsnProArg 321
Db 1171 TATGATCAGCTCATGGAGGATGGGAGAAAAAGTGACAGAAATAGAAAAATATCTCTCG 1230
QY 322 ArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArg 341
Db 1231 AGAAAGCTAAGAAAGCAAAACAGGGAATACTATGAAAGCAAGTTTCCAGAAATTCGA 1290
QY 342 LysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeu 361
Db 1291 AAACAAGAGACACAGCAAGAAAGATTTTCAG--CGAGTGGCGAGAGGGAGCTGCTT 1347
QY 362 SerMetSerAlaAlaArgSerGluHisGluValSerGluIleLeuLeuSerGlyLeu 381
Db 1348 TCAGCCACCATTTGAGAGTGAGCATGAGATTTCTGAAATATTATGATGGCTCTCTGAG 1407
QY 382 GlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAsp 401
Db 1408 CAGAGAAATATGAGAAACAAATGCGGCAGCTCTCTGTGATTCACCTATGATGTTGAT 1467
QY 402 AlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLys 421
Db 1468 GCAGAACAGACGACGTCAGTTTCAATTAACATGAATGGCTTATGGAGACCTATGAAA 1527
QY 422 ValTyrLysAspArgGlnValMetAsnMetTrpSerGlnGlnGluLysGluThrPheArg 441
Db 1528 GTGTATAAGATAGGCGAGTTTATGAATGTTTGGACTGACCATGAAAGAGAGATCTTTAAG 1587
QY 442 GluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLys 461
Db 1588 GACAAGTTTATCCAGCATCCAAAAAATTTGGACTAATTCATCATATTTGGAGAGGAAG 1647
QY 462 ThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSer 481
Db 1648 AGTGTCTCTGATGTGTTTGTATTACTATTATTAACCAAGAAAAATGGAATTTAAGCC 1707
QY 482 LeuValArgSerTyr---ArgArgArgGlyLysSerGlnGlnGlnGlnGlnGln 500
Db 1708 CTCCTCAGAGGAATATGAGAAACGAGAGGAGAAACAGCAAAATTTGTCGACCTCG 1767
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluLys 520
Db 1768 CAAGAGAAAAAGTAGAGAAAAAGAGAG-----GATAAAGCAGAAAAAAGAGAAAA 1821
QY 521 AspGluLysGluLysGluLysGluAlaGluLysGluLysProGluValGluAsn 540
Db 1822 AAGAGAGAAAAAGAGATGAGAGAAAAAGATGAAAGAAAGACTCAAGAAAT 1881
QY 541 AspLysGluAspLeuLeuLysGluLysThrAspThrSerGlyGluAspAspGlu 560
Db 1882 ACCAAGGAA-----AAGGACAAGATAGATGGTACAGCA--GAAGAAACTGAGGAA 1929
QY 561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGly 580
Db 1930 AGAGAGCAAGCCACACCCCGGGGGCGAAAGACTGCCAACAGTCAGGCGCCGCTAAGGGC 1989
QY 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
Db 1990 CGATCACCAGGTCCATGACAAACAGAGCTGAGCTGCCAGTCTGCGAGCCGCGGCT 2049
QY 601 Ser-----AlaGluLeuAlaSerMetGluLeuAsn 610
Db 2050 ACTGAAGAGCCCCACCATCTGCGCACCGCCACAGACCCATTTCTACAGAGCCTGTG 2109
QY 611 GluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuHis 630
Db 2110 GAGACCTCTCATGGAGAGAGAAATGGAAGTTGCTAAAAAAGGCTTAGTAGAACAT 2169
QY 631 GlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLys 650
Db 2170 GGTGCTAACTGGGCAATTTGTAATATGTTGGGAGCAAGAAAGTGAAGCTCAATGAAA 2229
QY 651 AsnPheTyrPheAsnTyrLysArgGlnAsnLeuAspGluIleLeuGlnHisLys 670

Db 2230 AACTTCTATTATTAACATATAAAGCGCACACAATCTTTGACAACCTCTTTACAGCAGATAAA 2289
QY 671 LeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysAlaProAlaAlaAla 690
Db 2290 CAGAAAACTTCAGAAAAACCTCGTGAAGACGAGATGTCTCAATGTGAAGTGTGCT 2349
QY 691 SerGluGluAlaAlaPheProProValValGluAspGluMetGluAlaSerGlyVal 710
Db 2350 TCCACTGTTTCTGCT-----CAGGAGGATGAAGATATTGAAGCTCC----- 2391
QY 711 SerGlyAsnGluGluGluMetValGluGluAlaLeuHisAlaSerGlyAsnGlu 730
Db 2392 -----AATGAAGAAGAAAAATCCAGAAGACAGCGAA----- 2421
QY 731 ValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSer 750
Db 2421 ----- 2421
QY 751 IleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProPro 770
Db 2422 -----GTTGAAGCTGTCAAG----- 2436
QY 771 AlaThrLeuGlyAlaAspGlyProProGlyProProGlyProProProProProPro 790
Db 2437 -----CCAGCGAGACAGCTCTGAAAAATGCTACTTCT 2469
QY 791 ArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProPro 810
Db 2470 CGAGAAACACAGAACCTCGGTTGAGTTGAG-----CCACACCGGAAACT 2517
QY 811 AlaPro---ProSerProSer---AlaProProProValValProLysGluLysGlu 828
Db 2518 GCACCCAGTACATCTCTCTCTTAGCAGTTCCAAGTACAAAAACCGCTGAAGTGAAGT 2577
QY 829 GluGluThrAlaAlaProProVal----- 837
Db 2578 GTGAGACCCAGGTGAATGACATCATGCTGCTGAGACAGCAGACAGATGGATGTAGAT 2637
QY 838 -----GluGluGlyGluGluGlnLysProProAlaAlaGluGluLeu 851
Db 2638 CAGCAGAGACAGTCTGAAGAGGGTCTGTTTGTGATCCCCCAGCTACCAAA--- 2694
QY 852 AlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGlu 871
Db 2695 ---GCTGACTCTGTGACGCTTGAAGTGAAGGTGCCAGAAAAACCATGCATCTAAAGTTGAA 2751
QY 872 GluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeu 891
Db 2752 GGTGATAATACCAAGAAAGAGACTTTGGATAGAGCC----- 2787
QY 892 LysAlaGluLysLysGluGlySerGlyArgAlaThrThrAlaLysSerSerGlyAla 911
Db 2788 ---AGTGAAGAGGTGAACCTAGAGATGAAGATTTGGTGTGATCAGCAAAATAATATGCC 2844
QY 912 -----ProGln---AspSerAspSerSerAlaThrCysSerAlaAspGluVal 926
Db 2845 CAAAGCCCGAGCCCGAGTCAGACATGATTCAGTGCACGTCGAGCGCTGATGAG--- 2901
QY 927 AspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerPro-----ArgProSer 943
Db 2902 ---GATGTGATGAGAGCGCAGAGCGCAGAGAAATGTTTCTATGAGACTCAAGACCTTCA 2958
QY 944 LeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLys-----ProLeu 961
Db 2959 CTGTAAACCCCACTGGATCTTACTCTGTC---TCATCTCCGTTAAACCAAAATCCACTG 3015
QY 962 AspLeuLysGlnLeuLysGlnArgAlaAlaIleProProIle----- 976
Db 3016 GATCTGCCACAGCTTCAGCATCGAGCTGTGTTATCCCACTATGTTCTCTGCACCCCA 3075
QY 977 -----GlnValThr 979
:::

Db 3076 TGTAAATACCAATTTGGAACCCAGTGCAGCGGTATGCTCTCTACAGCGACACATTTAA 3135
Qy 980 LysValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaPro 999
Db 3136 GCAATGCATGAGTCAGCACTCTCTGGAG----- 3162
Qy 1000 ProProGluAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerPro 1019
Db 3163 ---GAGCAGCGGAGAGACAGACAGATAGATTTGGAAATGATAGAGTTCTCAAGTCCA 3219
Qy 1020 ArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAla 1039
Db 3220 TGTGGCAGATCCAGAGTCCA-----AACAGAGAG----- 3249
Qy 1040 GlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValProPro 1059
Db 3250 -----TGG----- 3252
Qy 1060 ArgGluValLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaPro 1079
Db 3253 ---GAAGTCTCTCAGCGCTCTCACAT-----CAATTGATACTATCTCCCT 3297
Qy 1080 ProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgPro 1099
Db 3298 GAAGCGTTCGGCTTCOG-----ACAACCGACCAACAGCGCCCGCC 3342
Qy 1100 ProThrLysSerAsnProProProLeuLysSerSerAlaLysHisProSerValLeuGlu 1119
Db 3343 CTTCTCATCCCGTCATCCAAACACACAGTGGCTTCAGAAAA---CCATCTTTTATA--- 3396
Qy 1120 ArgGluLysAlaLysSerGlnGly-----MetSerValGlnLeuHis 1134
Db 3397 ---ATGGAGGCTCCATCTCACAGGGAACACAGGCACTATTGTACTCTCATATCAG 3453
Qy 1135 ValProTyrSerGluHisAlaLysAlaPro---ValGlyProValThrMetGlyLeuPro 1153
Db 3454 GCTTCTTACACTCAAGAAACACCAAGCGCTCAGTAGGACTATCTCTCTGGACTGCCA 3513
Qy 1154 LeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGlnLeuSer 1173
Db 3514 CGCAACAGGAATCTGCCAAATCAGCTACTTTCCTTACATCAAGCAGGAAGAAATTTCT 3573
Qy 1174 ProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSer 1193
Db 3574 CCGCAAGCCAAACTCACAACCTGAGGCTCTGTGGTC---AGGGCCCAACATGAAGT 3630
Qy 1194 ValLeuArgGlyThrAlaLeuGlySerValProGlySerLeuThrLysGlyLeuPro 1213
Db 3631 GTAGTCAGAGTACCGCA---GGAGCCATACAGAGGAAGTATACTCGGGAACTCCA 3687
Qy 1214 SerThrArgValProSerAspSerAlaLysThrTyrArgGlySerLeuHisGlyThr 1233
Db 3688 ACCAGCAAAATTTTCAGTGGAGAGCATCTCCATCCCTACCGGGCTCTATCACTCAGGGCACC 3747
Qy 1234 ProAla-----AspValLeuTyrLysGlyThrLeuThrArg 1245
Db 3748 CCGGCTCTGCCCGACAGTGCATACCAACAGAGCCTTGTGAGGGTCCATTCGAGA 3807
Qy 1246 IleLeuGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLys 1265
Db 3808 ATGCCCATTTGAAGACAGCAGTCTCT-----GAGAAAGGAGAGAGAAAGTGCATCCAA 3861
Qy 1266 GlyHisValLysLeuGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSer 1285
Db 3862 GGCCATGTTATTTATGAAGCAAAAGTGGACATATCTTGTATATGATATATATTAAGAT 3921
Qy 1286 ValThrGlnCysSerLysGluAspGlyArgSerSerGlyProProHisGluThrAla 1305
Db 3922 GCC-----CGAAGAGGACTAGGAGTCCAGAACAGCTCATGAATCAGT 3966
Qy 1306 AlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaLysSer----- 1323
Db 3967 TTA---AAGAGAGCTATGAATCAGTGGAGGAATATTAAGCAAGGATGTCAATGAGG 4023

Qy 1324 -----SerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArg 1339
Db 4024 GAGTCTCTGTATCAGCACCTGTAGAGGGGTGATATGCCAGCATTAACC-----AGG 4077
Qy 1340 HisSerProHis---HisLeuLysGluGlnHisIleArgGlySerIleThrGlnGly 1358
Db 4078 GGGAGTCTCTCATTCTGACCTCAAAAGAAAGAGTGTATTGTCTGGCTCCATAATGACGGG 4137
Qy 1359 IleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgGluAlaLysLeuLeu 1378
Db 4138 ACACCAAGAGCAACAACTGAAAGCTTTGAAGATGGCCTT---AAATATCCCAACAAAT 4194
Qy 1379 LysArgGluGlyThrProProProProProSerArgAspLeuThrGluAlaTyrLys 1398
Db 4195 AAAGGGAAAGT-----CCTCCATACAGCATTTGAAGGTGCCATT--- 4236
Qy 1399 ThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrVal 1418
Db 4237 -----ACCAAGGAAACCA---TATGATGGC---ATCACCACCATC 4272
Qy 1419 LysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeu----- 1433
Db 4273 AAAGAAATGGGGCGTTCATTTCATGAGATTCAGAGCAAGATATTTTAACTCAGAAAGT 4332
Qy 1434 ArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGln 1453
Db 4333 CGAAATCTCCAGAGTGGTCCAGAGCACACGGCCGATAATTGAGGGTTCATTTCACAG 4392
Qy 1454 GlyThrProLeuLysTyrAspThrGlyAlaSerThrGlySerLysLysHisAspVal 1473
Db 4393 GGCACACCAATAAAGTTTGACAAAC---TCAGGTCAATCTGCCATCAACACACATGTC 4449
Qy 1474 ArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMet 1493
Db 4450 AAATCTTAATCAGCGGCTTAGCAAACTATCCCTGGAAATCCCTCGCTGGAAATGTG 4509
Qy 1494 AlaAsp---AlaArgAlaLeuGluArgAlaCysTyrGlu-----GluSer 1507
Db 4510 CCAGAGACATAAAGTGTAGAACGGGAAATATGAGATGTGAAAGCAGCGCAGACC 4569
Qy 1508 LeuLysSerArgProGlyThrAlaSerSerSerGlySerIleAlaArgGlyAlaPro 1527
Db 4570 GTGCGTTCGGGCACACACAGTCAGTGTAAAGTCTGCGCCCTCTTGTATGATGACACAGT 4626
Qy 1528 ValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGly 1547
Db 4627 -----CTGCATGAAGCTCCCAAGACACACTGAGCCCTGGGATTTATGATGACACAGT 4680
Qy 1548 Ala-----ProPheAlaGlyHisLeuProArgGlySerProValThrMet 1562
Db 4681 GCACGGAGGAGCCCTGTGAGTTATCAAAACACCATGTCCAGAGGCTCACCCATGATGAAC 4740
Qy 1563 ArgGluProThrProArgLeuGlnGlySerLeuSerSerSerLysAlaSerGlnAsp 1582
Db 4741 AGAACTTCTGATGTTACAAATTCCT-----CCTAAACAAGTCTACCAATCATGAAC 4788
Qy 1583 ArgLeuLeuThrSerThrProArgGlu-----IleAlaLysSerProHisSer 1598
Db 4789 AGAAATCCGACTGACCTCCCTACCCAGGGAAAGTATCCAGCGAAGTCTCCAGTGCCT 4848
Qy 1599 ThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyVal 1618
Db 4849 GGGGTGGACCTGTCTGTGAGCCAC-----AGTCCGTTTGTATCCCATCACAGAGGACG 4902
Qy 1619 SerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIlePro 1638
Db 4903 ACTCAGGCGAGGTTTTATTGGAGCCACCTGCCACCAATTTGATCCA---GCCATGCTCT 4959
Qy 1639 ArgGlyIleProLeuAsp---AlaAlaAlaTyrTyrLeuProArgHisLeuAlaPro 1657
Db 4960 TTTCAGAGGCTTTGGATCTCGAGCGGTGCTTACTCTGTTTTCAGAGACAGCTTTCCACCA 5019

QY 1658 AsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAla 1677
Db 5020 ACTCCAGTTACCCAGTTCAGTATCAGCTTTAC----- 5052
QY 1678 AlaLeuGluAsn---ArgGlnThrIleLeuAsnAspTyrIleThrSerGlnMetHis 1696
Db 5053 GCATGGAGAACACAGACAGACAATCTTAAATGATTACATTACCTCACACAGATGCCAA 5112
QY 1697 HisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArg 1716
Db 5113 GTGACTTG-----CGTCCAGATGTGCCAGAGGACTCTCCCAAGA 5154
QY 1717 GluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleLeuAspLeuSerGln 1736
Db 5155 GAGCAGCACTGGGTCTCCCATACCCAGCA---ACGAGAGGAATCAATGACTGACCAAT 5211
QY 1737 ValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAsp 1756
Db 5212 ATGCCCTCCA---ACAAATTTTAGTGCCTCATCCAGGGGGAACAAGCACTCTCCCATGGAC 5268
QY 1757 ArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArg---HisSerSerSer 1775
Db 5269 AGAATCACTTATCTCTGTGCACAGATTACTTTCCTCCAGGCGGTACAACTCTGCT 5328
QY 1776 ProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGlu 1795
Db 5329 TCCATGTCTCCAGACACCCCAACACACCTT-----GCAGCTGTGCAAGTGTGAG 5379
QY 1796 ArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeu 1815
Db 5380 AGGGAACGGGAACGGGAGCGGAGAGGAGCGGAGCGGGAACGG-----ATTGCT 5430
QY 1816 ThrSerThrThrValGluHisAlaProIleThrArgProGlyThrGluGlnSerSer 1835
Db 5431 GCAGTCTCTCCACCTC-----TACTGCGGCGAGGCTCAGAACAG----- 5472
QY 1836 GlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAlaSerHisSer 1855
Db 5473 -----CCTGGCGGAGCTGGCAGTCATGGA 5496
QY 1856 HisAlaHisGlnHisSerProIleSerProArgThrGlnAspAla---LeuGlnGlnArg 1874
Db 5497 TATGTTGGCTCCCTCTCCCT---TCAGTAAGAACTCAGAGACCAATGTTGCAACAGAGA 5553
QY 1875 ProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLys 1894
Db 5554 CCAGTGTGTTTCCAGGAACCAATGGNACCAAGTGTATACACCTTTGGATCCACTGCT 5613
QY 1895 ProThrValLeuArgSerThrSerThrSerSerProVal-----ArgProAla 1910
Db 5614 CAGCTACGAATCATGCCACTGCTGCTGGGGGCCCTTCAATAAGCCAAAGGCGCTGCCAGCC 5673
QY 1911 AlaThrPheProAlaThrHisCysProLeuGlyThrLeuAspGlyValTyrPro 1930
Db 5674 TCCCGTTACAACTGCTCGGATGCC---CTGGCTGCTCTTGTGGATGCTGCAGCTTCT 5730
QY 1931 ThrLeuMetGluProValLeuLeuProLys-----GluAlaProArgValAla 1946
Db 5731 GCACCCAGATGATGTGTCCAAAACAAAGAGAGTAAAGCATGAAGCTGCAGGTTAGAA 5790
QY 1947 ArgProGluArgProArgAlaAsp----- 1954
Db 5791 GAAAAATTTGAGAGCAGGTCAGCAGCAGGTTAGTGAACAGCAGCAGCTAGAGCAAAAACC 5850
QY 1955 -----ThrGlyHisAlaPheLeuAlaLys 1962
Db 5851 CTGGAGTGGAGAGATCTGTTCACTGTTTATACACTTCTTCAGCGCTTTTCAAGTGGC 5910
QY 1963 ProProAlaArgSerGlyLeuGluProAlaSerSer-----ProSerLysGlySer 1979
Db 5911 AAGCCCC-----CAGCCTCATCTTTCAGTAGTTATCTTCTGAGGCTGGGAAA 5955
QY 1980 GluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAla 1999

Db 5956 GATAAAGGCGCT-----CCTCCAAAATCCAGATATGAGGAAGAGCTAAGACCAGAGGG 6009
QY 2000 LysAsn-----LeuAlaProHisHisAlaSerProAspProAlaProProAlaSer 2017
Db 6010 AAGACTACCATTTACTGCTAGCTAACTTCATAGACGTGATCATCCCGGCAAAATGCTCG 6069
QY 2018 AlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGlu 2037
Db 6070 GACAAGGATGCGAGGGAACGTGCTCTCAAAGTTACAGACTCTTCT----- 6114
QY 2038 LeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSer 2057
Db 6115 ---AGTAGCTTATCTTCTCACAGGTATGAACACCTAGCGATGCTATTAGGTGATAAGT 6171
QY 2058 ProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluLeu 2077
Db 6172 CTGCGCAGCTCACCTGCGCACCCAGGAGAAACTGCGACCTATCAGCCAGAGGTTGTT 6231
QY 2078 AspLysSerHisLeuGlyGluLeuArgProLysGlnProGlyProValLysLeuGly 2097
Db 6232 AAGGCAAAATCAAGCGGAAAATGATCTACCAGACAATATGAAGGACCA----- 6279
QY 2098 GlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSer 2117
Db 6280 -----TTACATCACTATCGACCA-----CAGCAGGAATCACCATCT 6315
QY 2118 ProLeuLeuGlnThrAlaProGly-----ValLysGlyHis 2129
Db 6316 CCCCAACACAGCTGCCCCCTTCTCAGGCGAGAGGAATGGGCAAGTGGCCAGGACC 6375
QY 2130 GlnArgValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArg 2149
Db 6376 CATCGCTGATCACACTGCTGATCATCTGTCAAATATCACACAAGATTTTGCTAGA 6435
QY 2150 HisHis-----ProGlnGlnLeuSerAlaProLeuProAlaProLeuTyr 2164
Db 6436 AATCAAGATTTCTCGCAGACTCCCGCAGCAG-----CCTCTACTTCTACATTCAG 6486
QY 2165 SerPheProGlyAla-----SerCysProValLeuAspLeuArgArgProProSerAsp 2182
Db 6487 AACTCACCTTCTGCTTGTGTTATCTACACCTGTG-----AGGACTAAACATCAAC 6537
QY 2183 LeuTyrLeuProProProAsp-----HisGlyAlaPro---AlaArgGly 2196
Db 6538 CGTTACAGCCAGAAATCCCGAGGCTCAGTCTGTCCATCATCAAGACCAGGTTCAGGGTC 6597
QY 2197 SerProHis-----SerGluGlyLysArgSerProGluProAsnLys 2211
Db 6598 TCTCCAGAAAAATCTTGTGGACAAATCCAGGGGAAGTAGGCTTGAAAAATCCCGAGAGG 6657
QY 2212 ThrSerValLeuGlyGlyGluAspGlyIleGluProValSerProProGluGlyMet 2231
Db 6658 AGTCAGTC-----TCTCCGAGCCCTACAGCCCATCTCCCGACCCAG----- 6702
QY 2232 ThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGln 2251
Db 6703 GTTCCGTTGTGTCATGAGAAACAGGACAGCTGCTGCTCTTGTCTCAGAGGGCGCAGAG 6762
QY 2252 ThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhe 2271
Db 6763 CTGCGAGAGCAGAGGAATGATGCCCGCTCACAGGGAGTATAAGCTACTTTCCTTCATTC 6822
QY 2272 PheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsn 2291
Db 6823 TTCACCAAGCTT---GAAATATCATCCCATGTTAAATCAAGAAAGCAGGAGATTTT 6879
QY 2292 LysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThr 2311
Db 6880 CGTAAGTTGAACCTCTCTGTTGGAGGTGACTCTGATATGTCAGCTGCTCAGCCAGGAAT 6939
QY 2312 GluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAla 2331

QY	791	ArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProPro	810
DB	2470	CGAGGAAACACAGAACTCGGGTGAAGTTGAG-----CCACACGAGAAACT	2517
QY	811	AlaPro---ProSerProSer---AlaProProValValProLysGluGluLysGlu	828
DB	2518	GCACCCAGTACATCTCCCTTAGCAGTTCCAAGTACAAAACACAGCTGAAGATGAAAGT	2577
QY	829	GluGluThrAlaAlaAlaProProVal-----	837
DB	2578	GTGGAGACCAGGTGAATGACAGCATCTAGTGTGAGACGACGACGAGATGGATGTAGAT	2637
QY	838	-----GluGluGlyGluGluGlnLysProProAlaAlaGluGluLeu	851
DB	2638	CACGAGGACACAGTGTCTGNAGAGGGTTCTGTTGTGATCCCCACCCGCTACCAAA---	2694
QY	852	AlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGlu	871
DB	2695	---GCTGACTCTGTGGACGTTGAAGTCAGGGTGCCAGAAAACCATGTCATCTAAAGTTGAA	2751
QY	872	GluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeu	891
DB	2752	GGTGATTAATACCAAGAAAGAGACTTGGATAGAGCC-----	2787
QY	892	LysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAla	911
DB	2788	---AGTGAGAAAGGTGGAACCTAGAGATGAAGATTTGGTGGTAGCTCAGCAATAAATGCC	2844
QY	912	-----ProGln---AspSerAspSerSerAlaThrCysSerAlaAspGluVal	926
DB	2845	CAAAAGCCCGAGCCCACTAGTCAGCAATGATTCAGTGCCACGTGCACGCGTGAATGAG---	2901
QY	927	AspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerPro-----ArgProSer	943
DB	2902	---GATGTGGATGGAGCCAGAGCCAGAGCAGAGATGTTCTATGGACTCAAGCCTTCA	2958
QY	944	LeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLys-----ProLeu	961
DB	2959	CTGTAAACCCCACTGGATCTATACTCGTC---TCATCTCGGTTAAACCAATCCACTCG	3015
QY	962	AspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProlile-----	976
DB	3016	GATCTGCCACAGCTTCAGATCGAGCTGCTGTATCCACCAATGTATCTCGACCCCA	3075
QY	977	-----GlnValThr	979
DB	3076	TGTAACATACCAATTGGAAACCCCACTGAGCGGTATGCTCTACCAGCCACATATAA	3135
QY	980	LysValHisGluProProArgGluAspAlaAlaProThrLysProAlaProAlaPro	999
DB	3136	GCAATGTGATGAGTCAGCACTCCTCGAG-----	3162
QY	1000	ProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerPro	1019
DB	3163	---GAGCAGCCGCAGACACAGACAGATAGATTGGAAATGTAGAAGTTCTTCAAGTCCA	3219
QY	1020	ArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAla	1039
DB	3220	TGTGGCACATCCAAGAGTCCA-----AACAGAGAG-----	3249
QY	1040	GlnLysLeuProGlyAspProProCysThrPThrSerGlyLeuProPheProValProPro	1059
DB	3250	-----TGG-----	3252
QY	1060	ArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaPro	1079
DB	3253	---GAAGTCCITCAGCCTGCTCCACAT-----CAATTGATAACTTAATCTCCCT	3297
QY	1080	ProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgPro	1099
DB	3298	GAAAGCGGTTCCGGCTTCGG-----ACAACTGACCACCAACGCGCCACCGGCC	3342

1100 ProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuGlu 1119
 DB CTTCTCATCCCGTCTCCAAAACACAGTGGCTTCAGAAAA---CCATCTTTTATA--- 3396
 1120 ArgGlnIleGlyAlaIleSerGlnGly-----MetSerValGlnLeuHis 1134
 DB ATGGGAGGCTCCCATCTCACAGGGACACACAGGCACCTTATTGACTTCTCATAAATCAG 3453
 1135 ValProTyrSerGlnHisAlaLysAlaPro---ValGlyProValThrMetGlyLeuPro 1153
 DB GCTTCTCCTACACTCAAGAAACACCAAGCGCTCAGTAGGATCTATCTCTCTGGACTGCCA 3513
 1154 LeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGlnLeuSer 1173
 DB CGGCAACAGGAATCTCCAAATCAGCTACTTTGCCCTACATCAGCAGGAAGAAATTTCT 3573
 1174 ProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSer 1193
 DB CCCCGAAGCCAAAATCACAACCTGAGGGCTGTGGTCT---AGGGCCCAACATGAAGT 3633
 1194 ValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyLeuPro 1213
 DB GTAAGTCAGAGGTACCGCA---GGAGCCATACAGAAGGAAGTATTAATCTGGGGAATCCA 3687
 1214 SerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThr 1233
 DB ACAGCAAAATTTTCAGTGGAGAGCATTCCTCCCTACGGGGCTCTATCACTCAGGCGACC 3747
 1234 ProAla-----AspValLeuTyrLysGlyThrIleThrArg 1245
 DB CCGGCTCTGCCCCAGACTGGCATACCAACAGAGGCTTTGGTGAAGGGTGCATATTCGAGA 3807
 1246 IleIleGlyGluAspSerProSerArgLeuAspArgIleValGlyGluAspSerLeuProLys 1265
 DB ATGCCCATTTGAAGACACAGCAGTCTCT-----GAGAAAGGCAGAGAGGAGCTGCATCCAAA 3861
 1266 GlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSer 1285
 DB GGCCATGTTTATGAAGGCAAAAGTGGACATATCTTGTCATATGATATTAAGAAAT 3921
 1286 ValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGluThrAla 1305
 DB GCC-----CGAGAAGGAGCTAGAGTCCAGAACAGCTCATGAATCAGT 3966
 1306 AlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSer----- 1323
 DB TTA---AAGAGAAGCTATGAATCAGTGGAAAGAAATATAAAGCAAGGATGTCAATCAGG 4023
 1324 -----SerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArg 1339
 DB GAGTCTCTGTATCAGCACCGTGTAGAGGGCTGATATGCCAGCATTTACCC-----AGG 4077
 1340 HisSerProHis---HisLeuLysGlnGlnHisIleArgGlySerIleThrGlnGly 1358
 DB GGGAGTCTCTATTCTGACCTCAAAGAAAGGACTGTATTGTCTGGCTCCATAATATCGAGGGG 4137
 1359 IleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeu 1378
 DB ACACACAGAGCAACAACCTGAAGCTTTGAAGATGGCCCT---AAATATCCCAACAAATT 4194
 1379 LysArgGluGlyThrProProProProProSerArgAspLeuThrGluAlaTyrLys 1398
 DB AAAGGGAAAGT-----CCTCCCATACGACATTTGAAGGTGCCATT--- 4236
 1399 ThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrVal 1418
 DB -----ACCAAAAGGAAAAACCA---TATGATGCC---ATCACCCACCATC 4272
 1419 LysGluAlaGlyArgSerIleHisIleThrProArgGluGluLeu----- 1433
 DB AAAGAAATATGGGCGCTTCCATTCATGAGATTCAGAGCAAGATATTTAACTCAGGAAAGT 4332
 1434 ArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGln 1453

Db	4333	CGAAAACTCCAGAAAGTGGTCCAGAGACACACGGCGGATAAATTGAGGGTTCATTTTCCACG	4399
Qy	1454	GlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspVal	1473
Db	4393	GGCACACCAATAAAGTTTGACACAAC--TCAGGTCAATCTGCCATCAACACATATGC	4449
Qy	1474	ArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMet	1493
Db	4450	AAATCCTTAATCACGGGGCTAGCAAACTATCCCGTGGAAATGCCTCGCTGGAAATTGTG	4509
Qy	1494	AlaAsp--AlaArgAlaLeuGluArgAlaCysTyrGlu-----GluSer	1507
Db	4510	CCAGAGAACATATAAAGTGGTAGAACGGGGAATAATGAGATGTGAAAGCAGCGCAGACC	4569
Qy	1508	LeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaPro	1527
Db	4570	GTGGGTCCCGGCACACGTCAGTGGTAGCTCTGGCCCTCCGTTCTTAGTCCACA---	4626
Qy	1528	ValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGly	1547
Db	4627	-----CTGCATGAAGTCTCCAAAGCACAACTGAGCCCTGGGATTTATGATGACACCAGT	4680
Qy	1548	Ala-----ProPheAlaGlyHisLeuProArgGlySerProValThrMet	1562
Db	4681	GCACGGAGGACCCCTGTGAGTTATCAAAACACCACTGTCAGAGCTCACCCTATGATGAAC	4740
Qy	1563	ArgGluProThrProArgLeuGlnGlySerLeuSerSerSerLysAlaSerGlnAsp	1582
Db	4741	AGAACTCTGATGTACAAATTCCT-----CCTAACAAAGTCTACCAATCATGAA	4788
Qy	1583	ArgLysLeuThrSerThrProArgGlu-----IleAlaLysSerProHisSer	1598
Db	4789	AGGAAATCGACACTGACCCCTACCAGAGGGAAAGTATCCAGCGGAAGTCTCCAGTGCCT	4848
Qy	1599	ThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyVal	1618
Db	4849	GGGGTGGACCCCTGTCGTGAGCCAC-----AGTCCGTTTGCATCCCATCACAGGCGAGC	4902
Qy	1619	SerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIlePro	1638
Db	4903	ACTCGAGCGAGGTTTATGAGCGCACCTGCGCCACGCAATGGATCCA---GCCATGCCT	4959
Qy	1639	ArgGlyIleProLeuAsp---AlaAlaAlaIleTyrTyrLeuProArgHisLeuAlaPro	1657
Db	4960	TTTCACAGGGCTTTGGATCTCGACGGCTGCTTACCTGTTTCAGACAGACGTTTCACCA	5019
Qy	1658	AsnProThrTyrProHisLeuTyrProProTyrIleuIleArgGlyTyrProAspThrAla	1677
Db	5020	ACTCCAGGTTACCCAAGTCAGTATCAGCTTTAC-----	5052
Qy	1678	AlaLeuGluAsn---ArgGlnThrIleIleAsnAspTyrIleThrSerGlnMetHis	1696
Db	5053	GCAATGAGAAACACAGACAGCAATCTTAATGATTACATTACTCACACAGATGCAA	5112
Qy	1697	HisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArg	1716
Db	5113	GTGAACCTTG-----CGTCCAGATGTGGCCAGAGGACTCTCCCAAGA	5154
Qy	1717	GluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGln	1736
Db	5155	GAGCAGCCACTGGGTCTCCCATATCCACGACA---ACGAGAGGAATCATTTACCTGACCAAT	5211
Qy	1737	ValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAsp	1756
Db	5212	ATGGCTCCA---ACAAATTTTAGTGCCTCATCCAGGGGGAACAGACATCTCTCCCATGGAC	5268
Qy	1757	ArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArg---HisSerSerSer	1775
Db	5269	AGAACTACTATATCTCTGGTACACAGATTACTTTCCCTCCCGCGGTACAACTCTGCT	5328
Qy	1776	ProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGlu	1795

D	b	5329	TCCATGCTC	CAGGACACC	CAACACCTT-----GCAGCTGCTCGAAGTGCTGAG	5379
Q	y	1796	ArgGluArgAsp	ArgGluArgAsp	ArgGluArgGluLysSerIleLeu	1815
D	b	5380	AGGGNACGGG	NACGGGAGCGG	GAGCGGGACCG-----ATTGCT	5430
Q	y	1816	ThrSerThrThr	ValGluHisAlaPro	IleTrpArgProGlyThrGluInSerSer	1835
D	b	5431	GCAGGTTCCT	CCGACCTC-----TACCTGGCGGCAGCTCAGAACAG-	5472	
Q	y	1836	GlySerSerGly	SerSerGlyGlyGlySerSerSer	ArgProAlaSerHisSer	1855
D	b	5473	-----	-----CCTGCCGCACCTGGCAGTCATGGA	5496	
Q	y	1856	HisAlaHisGln	HissPrProlleserPro	ArgThrGlnAspAla---LeuGlnGlnArg	1874
D	b	5497	TATGTTCCGCT	CCCCCT--TCAGTAAGA	AACTCAGGAGCACCATGTTGCCAACACAGA	5553
Q	y	1875	ProSerValLeu	HisAenThrGlyMet	GlyLylelleThrLaValGluProSerLys	1894
D	b	5554	CCCAGTGT	TTTTCCAAGGAACA	CAATAGGAACAGTGTAAATCACACTTTGGATCCAACATGCT	5613
Q	y	1895	ProThrValLeu	ArgSerThrSerThrSer	ProVal-----ArgProAla	1910
D	b	5614	CAGCTACGAAT	CATGCCACCTCGCTCG	GGGGCCCTTCAATAAGCACAAGCGCTCCGACGC	5673
Q	y	1911	AlaThrPhePro	ProAlaThrHisCys	ProLeuGlyGlyThrLeuAspGlyValTyPro	1930
D	b	5674	TCCCGTTACA	ACACTGCTGCGGATGC-CTGGCTGCTCTTG	TGGATGCTGCAGCTTCT	5730
Q	y	1931	ThrLeuMetGlu	ProValLeuLeuProLys-----GluAlaPro	ArgValAla	1946
D	b	5731	GCACCCAGCAT	GGATGTGTCCAAAACA	AAGAGAGTAGCATGAAGCTGCCAGGTTAGAA	5790
Q	y	1947	ArgProGluArg	ProArgAlaAsp-----	-----	1954
D	b	5791	GAATAATTGAG	AACAGCGTCAGCAGCAGT	AGTGAAACAGCAGCAGCTAGAGCAGACAAAAACC	5850
Q	y	1955	-----	-----ThrGlyHisAlaPheLeuAlaLys	1962	
D	b	5851	CTGAGGTGGAG	AGAGATCGTTTCAGTGTTTATACACTTCTTC	CAGCTTCCAGTGGC	5910
Q	y	1963	ProProAlaArg	SerGlyLeuGluProAlaSerSer-----ProSerLysGlySer	1979	
D	b	5911	AAGCCC-----	-----CAGCCTCATCTTCAGTAGTATTCTGAGCGCTGGGAAA	5955	
Q	y	1980	GluProArgPro	LeuValProProValSerClyHisAlaThrIleAlaArgThrProAla	1999	
D	b	5956	GATAAAGGGCCT-----	-----CCTCCAAAATCCAGATATCAGGAAGAGCTTAAGACACAGAGGG	6009	
Q	y	2000	LysAsn-----	-----LeuAlaProHisHisAlaSerProAspProProAlaProProAlaSer	2017	
D	b	6010	AAGACTACCATT	ACTGACGTAACTTCATGACGTGATCATC	CCCCGGCAAAATGGCTCG	6069
Q	y	2018	AlaSerAspPro	HisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGlu	2037	
D	b	6070	GACAAGGATGCG	AGGGAACGTGGCTCTCAAAGTTTCAGACTCTTCT	-----	6114
Q	y	2038	LeuArgSerLeu	GlyTyHisGlySerSerTySerProGluGlyValGluProValSer	2057	
D	b	6115	---AGTAGCTTAT	CTTCTCAGATGATGAACACCTACGCGATGCTAT	TATGAGTGATTAAGT	6171
Q	y	2058	ProValSerSer	ProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeu	2077	
D	b	6172	CTGCCAGCTCA	CTGCGCCACCCAGGAGNAACCTGACAGACCTAT	CAGCCAGAGTTGTT	6231
Q	y	2078	AspLysSerHis	LeuLeuGlyGlyGluLeuArgProLysGlnProGlyProValLysLeuGly	2097	
D	b	6232	AAGGCMAATCA	AGCGGAAATGATCTCTACAGACAATATGAAGGACCA	-----	6279
Q	y	2098	GlyGluAlaAla	HisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSer	2117	
D	b	6280	-----TTACATCA	CTACTTCGACCA-----CAGCAGGAATACCACTCT	6315	

```

QY 2118 ProLeuLeuGlnThrAlaProGly-----VallysGlyHis 2129
DB 6316 CCCCAACAACAGTGCCTCCCTTTCACAGGAGAGGGAATGGGCAAGTGCCTCCAGGACC 6375
QY 2130 GlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArg 2149
DB 6376 CATCGCTGATACACTTCTGATCATCATCTGTCAAAATATATACACAGATTTTCTAGTA 6435
QY 2150 HisHis-----ProGlnGlnLeuSerAlaProLeuProAlaProLeuTyr 2164
DB 6436 AATCAAGTTTCTCGCAGACTCCACAGCAG-----CCTCCTACTTCTACATTCAG 6486
QY 2165 SerPheProGlyAla-----SerCysProValLeuAspLeuArgProProSerAsp 2182
DB 6487 AACTCACTTCTGCTTTGTGATCTACACCTGTG-----AGGACTAAACATCAAC 6537
QY 2183 LeuTyrLeuProProAsp-----HisGlyAlaPro---AlaArgGly 2196
DB 6538 CGTTACAGCCCAATCCCGAGTCTAGTCTGTCATCATCAAGACCAAGGTTCAAGGTC 6597
QY 2197 SerProHis-----SerGluGlyGlyLysArgSerProGluProAsnLys 2211
DB 6598 TCTCCAGAAATCTGTGGACAATCCAGGGGAAGTAGGCTGGAAAAATCCCCAGAGAG 6657
QY 2212 ThrSerValLeuGlyGlyGluAspGlyIleGluProValSerProGluGlyMet 2231
DB 6658 AGTCACGTC-----TCTCCGAGCCCTACGAGCCCATCTCCACCCACAG----- 6702
QY 2232 ThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGln 2251
DB 6703 GTTCCGGTTGTGATGAGAAACAGGACAGCTGCTGCTGCTCAGAGGGCGCAGAG 6762
QY 2252 ThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhe 2271
DB 6763 CTTGACAGCAGAGGAATGATGCCGCTCACAGGGAGTATAGCTACTTGTGCTTCATTC 6822
QY 2272 PheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysGlnGluLeuAsn 2291
DB 6823 TTCAACCAAGTTT---GAAATCATATACCCATGTTTAAATCAAGAAGCAGGAGATTTT 6879
QY 2292 LysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnLysSerGlnProGlyThr 2311
DB 6880 CGTAAGTTGAATCCTCTGCTGGAGGTGACTCTGATATGCGAGCTGCTCAGCCAGGAAT 6939
QY 2312 GluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAla 2331
DB 6940 GAGATCTTTAATCTGCAGCAGTTACTACTGCTCAGGCTCAGTTAGTCTTAGAGGCCATTCT 6999
QY 2332 ValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMet 2351
DB 7000 TTTGCTGATCCTGCAGT---AATCTGGGCTGGAAGACATTATCAGGAAGCTCTCATG 7056
QY 2352 GlyLysTyrAspGlnTrpGluGlu-----SerProLeuSerAla 2365
DB 7057 GGAAGCTTTGATGACAAAGTTGAGGATCATGAGATTGCTATGCTCCAGCCTATGGGAGTA 7116
QY 2366 AsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAla 2385
DB 7117 -----GTGCTGTGATCTGCCAACACCTCAGTT-----GTGACCAAGT 7152
QY 2386 AlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGly-----LysAla 2403
DB 7153 GGTGAGACAGAGAGAGAGAGGAGGACCCATCATCTTTCAGGAGGAGTTTGCMAACCA 7212
QY 2404 LysValSerGlyArgProSerSerArgLysAlaLysSerProAlaPro-----GlyLeu 2421
DB 7213 AAGCTGATCAGCAAGTCAAAACAGCAGGAAATCTTAAGTCTCTTACTTACCTGGCGAGGCTAC 7272
QY 2422 AlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCysAsnArg 2441
DB 7273 TTAGAACGGAACGGCCCTCTTCTAGTCTCTCTGTATCATTCAGAGGGGATACCATAGG 7332

```

```

QY 2442 ArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrPro 2461
DB 7333 CAGAGGCCA-----GGTGGGCTGGGAAGACAGAGGCCCTCTTCAACAGGCTCAACTCAG 7386
QY 2462 PheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProPro 2481
DB 7387 TTTCTTATTAACCTCTGACTATGCGGATG-----CTCAGCAGTACTTCCACCAACA 7437
QY 2482 ProGlyLeuProAlaGlySerGly---ProLeuAlaGlyProHis-----HisAla 2497
DB 7438 CCGATTGATGCTGCTCCCTTCGCGTGAACCAACAGCAGCTCTCTCACCAAGACAGGATC 7497
QY 2498 TrpAspGluGluProLysProLeuLeuCysSerGlnTyrGluThrLeuSerAspSerGlu 2517
DB 7498 TGGAGCGAGAGCGCTGCCCACTGCTCTCAGCACAGTACGAGACCTTGTGCGATAGTGT 7557

RESULT 12
AF087856 8018 bp mRNA linear PRI 23-JUL-2003
LOCUS Homo sapiens nuclear receptor co-repressor mRNA, complete cds.
DEFINITION AF087856
ACCESSION AF087856
VERSION AF087856.1 GI:33150575
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Tu,Q., Yu,L., Dai,F., Yue,P., Huang,H. and Zhao,S.Y.
TITLE Cloning and characterization of a new human cDNA homologous to
JOURNAL murine nuclear receptor co-repressor mRNA
REFERENCE 2 (bases 1 to 8018)
AUTHORS Zhang,M., Yu,L., Zhou,Y., Hu,P.R., Xin,Y.R. and Zhao,S.Y.
DIRECT SUBMISSION
TITLE Submitted (27-AUG-1998) Lab of Human Gene Research, Institute of
JOURNAL Genetics, Fudan University, No. 220 Handan Rd., Shanghai 200433,
P.R. China
FEATURES
1..8018 Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
85..7116
/note="ncr"
/codon_start=1
/product="nuclear receptor co-repressor"
/protein_id="AAP97166.1"
/db_xref="GI:33150576"
/translation="MSSSVYPPNQAFSTQSRSPPHSVKVFPTSTTHHQDPAFGGKH
EAPSPISGQPCGDDQNASPKLSKEELIQMDVRDREIAKVEQILKKKQQQLLE
EAAKPEPEKVPSPPPVQKHSIVQIYYDENRKAEEAHKIFGLGPKVELPLYNQ
SDTKVYHENIKTNQVMRKLLILFKRRNHARKQREKICORYDQLMEEAKKVDRIEN
NPRKAKSKTEYVEKQPEIRKOREQOERFORVGORGAGLSATIASRSEHSIIDI
GLSESENEKQMRQLSVIPMMFDEAQRVFLNNGLMGDMPMKVYKOROFNMVMTDH
EKBIKPKRPIQHPKNGFLIASYLERSKSPDCVLYYLTCKENYKALVRNNGKGRGR
NOQIARSPQEEKEEDKAEKKEEKKDEEKKEDSKEDSKTEKKIKLDSTAE
ETEEEOATPRGRKTANSQGRKGRITRTSMNEAAAAAASAAAAATEEPPLPPEP
ISTEPVETSRTEEMEEVAKGLVEHGRNWAIAKXVGTKEAQCNFYFNKRRHNL
DNLLQHKQTKSRKPREERDVQSQSVASTVAQEDDIKASNEENDESEGSNS
DTESAPSPVEAVKPSPENATSRGTEPAVELEPTTAPSTPSLAVPTKPA
EDESVEQUNDSISATAEQMDVDQOEAEESGVCDDPPATKADSVDEVVRPNHA
SKVEGDKTERLDRAASKVEPRDELVVAQINAOQREPEQSDNDSSATCSADEVDG
EPEQRMPFMDSPKSLNPTGSLVSSPLKPNPLDLQLQHRAAVIFPMWSTPCNIP
ICTPQSGVALYQRHIAKMHESALLEQRQREIDLECRSTSPCGTSKSPNREVL
QPAHQVITNLPEGVRLPTTRTPPPLIPSSKTTVASKEKPSFMGSGISQGTPTGY
LTSHQNASYTOETPKPSVSGISLGLPROQESAKSATPLYIKQEEFSPRSQSQEGLL
VRAQHEGVRTGAGAIQESITRECTPSKISVESIPSLRGSITQGTALPQTGIPTEA
LVKGSISRNPIDEDSPFEKREARAAKSHVIEGKSGHILYDNIKNAREGTSPRTAH
EISLKRSEVSGNKKQGMSPRSPVAPLEGLICRALPRGPHDLKERTVLSGIM
RODILATTESFEDGLKPKQIKRESPPITRAFEGAITKGKPGDITIKEMGKRSIHIP
RODILATTESFEDGLKPKQIKRESPPITRAFEGAITKGKPGDITIKEMGKRSIHIP
ISQGTPIKFDNNNSQSAIKHNKSLITGPKL
SRGMPLIEIPENIKVVERGKVEDVKAGETVRSRHTSVSSGFSVLRLSTLHPAKQL

```


Db 3682 TCTGGCTCCATAATGTCAGGGGACACCAAGAGCAACAACCTGAAAGCTTTGAAGATGGCCCTT 3741
Qy 1372 ArgArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProSerArg 1391
Db 3742 ---AAATATCCCAACAAATTAAGGGAAGT-----CCTCCCATACGA 3783
Qy 1392 AspleuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHis 1411
Db 3784 GCATTGAAGTGCCATT-----ACCAAGGAAACCA---TAT 3819
Qy 1412 GluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGlu 1431
Db 3820 GATGCG---ATCACCAACCATCAAGAAATGGGGGCTTCCATTCAIGAGATTCCAAAGCAA 3876
Qy 1432 GluLeu-----ArgHisThrProGluLeuProLeuAlaProArgProLeu 1446
Db 3877 GATATTTTAACTCAGGAAAGTCGGAAGTCCAGAGTGGTCCAGAGCAGACGCGCCGATA 3936
Qy 1447 LysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThr 1466
Db 3937 ATTGAGGGTTCCATTTCACGGGCACACCAATAAAGTTTGACAACAAC---TCAGGTCAA 3993
Qy 1467 GlySerLysLysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProPro 1486
Db 3994 TCTGCCATCAACACAATGTCAAAATCTTAATACACGGGGCTAGCAAACTATCCCGTGGGA 4053
Qy 1487 ValHisProLeuAspValMetAlaAsp---AlaArgAlaLeuGluArgAlaCysTyrGlu 1505
Db 4054 ATGCTCTCGGTGGAATTTGTCAGAGNACATAAAGTGTAGAACGGGGAATATATGAG 4113
Qy 1506 -----GluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGly 1520
Db 4114 GATGTGAAGCAGCGAGACCGTGGTCCCGGCACACGTCAAGTGGTAAGCTCTGGCCCC 4173
Qy 1521 SerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerPro 1540
Db 4174 TCCGTTCTTAGTCCACA-----CTGCATGAAGCTCCCAAGAGCACAACACTGAGCCCT 4224
Qy 1541 LeuThrTyrGluAspHisGlyAla-----ProPheAlaGlyHisLeuPro 1555
Db 4225 GGGATTATGATGACACCATGTCAGGAGGACCCTGTGAGTTATCAAAACACCATGTCC 4284
Qy 1556 ArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlySerLeuSer 1575
Db 4285 AGAGGCTCACCATGATGAACAGA-----ACTTCTGATGTTCT 4323
Qy 1576 SerSerLysAla---SerGlnAspArgLysLeuThrSerThrProArgGlu----- 1591
Db 4324 TCTAACAGTCTCAACATCATGAAAGGAAATCGACACTGACCCCTACCCAGAGGGAAGT 4383
Qy 1592 ---IleAlaLysSerProHisSerThrValProGluHisHisProHisProIleSerPro 1610
Db 4384 ATCCAGGAAGTCTCCAGTCGCTGGGTGACCCCTGTCGTGAGCCAC-----AGTCCG 4437
Qy 1611 TyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeu 1630
Db 4438 TTTGATCCCATCATCAGAGGCGACACTGACGGCGAGGTTTATCGAGGCCACTGCCAAG 4497
Qy 1631 AlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyr 1650
Db 4498 CACTTGGATCCA---GCCATGCCCTTTTACAGGGCTTTGGATCTCGCGGCTGTACCTG 4554
Qy 1651 LeuProArgHisLeuAlaProAsnProThrTyrProHisHisLeuTyrProTyrLeuIle 1670
Db 4555 TTTCAGAGACAGCTTTTCAACAACTCCAGTTACCCAGTCAAGTCAAGTTCAGCTTTAC 4608
Qy 1671 ArgGlyTyrProAspThrAlaAlaLeuGln---ArgGlnThrIleIleAsnAspTyr 1689
Db 4609 -----GCAATGGAGAACAACAAGACACAGACACAATCTTAATGATTAC 4647
Qy 1690 IleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMet 1709
Db 4648 ATTACCTCACAAACAGATGCAAGTCACTTG-----CGTCCAGATGTG 4689

Qy 1710 LeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArg 1729
Db 4690 GCCAGAGGACTCTCCCAAGAGAGCAGCAGCTGGTCTCCCATACCCAGCA---ACGAGA 4746
Qy 1730 GlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGly 1749
Db 4747 GGAATCATTCACCTGACCAATATGCTCCA---ACAATTTTAGTGCCTCATCCAGGGGA 4803
Qy 1750 ThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSer 1769
Db 4804 ACAAGCACTCTCCCATGGACAGATCACTATATCTTGGTACACAGATTACTTTCCCT 4863
Qy 1770 SerArg---HisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysPro 1788
Db 4864 CCAGGCGGTACAACTCTGCTCTCCATGCTCTCCAGGACACCAACACACCTT----- 4914
Qy 1789 ThrThrThrSerSerGluArgGluArgAspArgGluArgAspArgAspArg 1808
Db 4915 GCAGCTCTCAAGTGTCTGAGAGGAAACGGGAGCGGAGCGGAGAGCGGAGCGG 4974
Qy 1809 GluArgGluLysSerIleLeuThrSerThrThrValGluHisAlaProIleTyrArg 1828
Db 4975 GAACGG-----ATTGCTGACGCTCTCCGACCTC-----TACTGCGG 5013
Qy 1829 ProGlyThrGluGlnSerSerGlySerSerGlyGlyGlyGlyGlySerSer 1848
Db 5014 CCAGCTCAGAACAG-----CCT 5031
Qy 1849 SerArgProAlaSerHisHisAlaHisGlnHisSerProIleSerProArgThrGln 1868
Db 5032 GGCGCAGCTGCGCATCATGATATGTTGCTGCCCTTCCCT---TCAGTAAGAACTCAG 5088
Qy 1869 AspAla---LeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIle 1887
Db 5089 GAGACCATGTTGCAACAGACCCAGTGTTCACAGAACCAATGGAACAGTGTATC 5148
Qy 1888 ThrAlaValGluProSerLysProThrValLeuLeuSerThrSerSerSerProVal 1907
Db 5149 ACACCTTTGGATCCCACTGCTCAGCTACGAATCATGCCCTGCTGCTGGGGCCCTTCA 5208
Qy 1908 -----ArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGly 1923
Db 5209 ATAAGCCAGGCTGCCAGGCTCCCGTTACAACTCTCGGATGCC---CTGGCTGCT 5265
Qy 1924 ThrLeuAspGlyValTyrProThrLeuMetGluProValLeuProLys----- 1940
Db 5266 CTTGTGATGCTGAGCTTCTGCACCCAGATGATGTGTCCAAACAAAGAGAGTAAG 5325
Qy 1941 ---GluAlaProArgValAlaArgProGluArgProArgAlaAsp----- 1954
Db 5326 CATGAAGCTGCCAGTTAGAAGAAATTTGAGAGAGCAGGTCAAGCAGCAGTTAGTGAACAG 5385
Qy 1955 -----Thr 1955
Db 5386 CAGCAGCTAGACGACGAGAAACCTCGAGGTGGAGAGAGATCTGTTCAAGTGTATACACT 5445
Qy 1956 GlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSer--- 1974
Db 5446 TCTTACGCTTCCAGTGGCAGGCC-----CAGCCTCATCTTCAGTA 5490
Qy 1975 -----ProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAla 1992
Db 5491 GTTTATTCTCAGGCTGGGAAAGATAAAGGCT-----CCTCCAAATCCAGATATGAG 5544
Qy 1993 ThrIleAlaArgThrProAlaLysAsn-----LeuAlaProHisHisAlaSerProAsp 2010
Db 5545 GAAGAGCTAAGGACGAGGAGGAACTACCATTAAGTCTGAGTAACTTTCATAGAGCTGATC 5604
Qy 2011 ProProAlaProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPro 2030
Db 5605 ATCACCCGGCAATGCTCTCGGACAGAGATGCGAGGAAACGTGGCTCTCTCAAAAGTTGACAC 5664

QY 2031 PheSerIleGlnGluLeuLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPro 2050
Db 5665 TCCTTCT-----AGTAGCTTATCTCTCACAGGTATGAACACACCTAGC 5706
QY 2051 GluGlyValGluProValSerProValSerProSerLeuThrHisAspLysGlyLeu 2070
Db 5707 GATGCTATTGAGTGATAAGTCTGCCAGCTCCTGCCAGCCACCCAGGAGAACTGCAG 5766
QY 2071 ProLysHisLeuGlnGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGln 2090
Db 5767 ACCTATCAGCAGGAGGTGTTAAGGCAATCAAGCGGAAATGATCTTACCAGACAATAT 5826
QY 2091 ProGlyProValLysLeuGlyGlyGluAlaHisLeuProHisLeuArgProLeuPro 2110
Db 5827 GAAGGACCA-----TTATCATCATATCGACCA----- 5853
QY 2111 GluSerGlnProSerSerProLeuLeuGlnThrAlaProGly----- 2125
Db 5854 ---CAGCAGGAATCACCATTCTCCCAACACAGCTGCCCTCTTTCACAGCGCAGGGA 5910
QY 2126 -----ValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluVal 2142
Db 5911 ATGGGGCAAGTGGCCAGGCCATCGGCTGATCACACTTCTGCTGATCATCTGTCAATT 5970
QY 2143 IleThrGlnAspTyrThrArgHisHis-----ProGlnGlnLeuSerAla 2157
Db 5971 ATACACAGAATTTTGTAGAAATCAAGTTTCTCGCAGACTCCCGCAGAG----- 6021
QY 2158 ProLeuProLalaProLeuTyrSerPheProGlyAla-----SerCysProValLeuAsp 2175
Db 6022 CCTCTACTTCTATCTCCAGAACTCACCTTCTGCTTGTGATATCATACCTGTG----- 6075
QY 2176 LeuArgArgProProSerAspLeuTyrLeuProProAsp-----His 2190
Db 6076 ---AGGACTAAACATCAAAACCGTTACAGCCCAAGATCCCGAGCTCAGTCTGTCCATCAT 6132
QY 2191 GlyAlaPro---AlaArgGlySerProHis-----SerGluGlyGlyLys 2204
Db 6133 CAAGACACAGTTTCAAGGGTCTCTCCAGAAATCTTGAGCAAAATCCAGGGGAGTAGG 6192
QY 2205 ArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGlyLysGlyLysGlyPro 2224
Db 6193 CCTGGAATAATCCCGCAGAGAGTACAGTCT-----TCTTGGAGCCCTACGAGCCC 6243
QY 2225 ValSerProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeu 2244
Db 6244 ATCTCCCAACCCAG-----GTTCCGGTGTGTGATGAGAAACAGCAGCAGCTTGCTGCTC 6297
QY 2245 LeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsn 2264
Db 6298 TTGTCTCAGAGGGCGCAGAGCTGCAGACAGAGGAATGATGCCCTCACCGAGGAGT 6357
QY 2265 ThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLys 2284
Db 6358 ATAAGCTACTTGCCTTCTTCCCAAGCTT---GAAATAATCATCACCCTGTTTAA 6414
QY 2285 SerLysLysGlnGluIleAsnLysLeuAsnThrHisAsnArgAsnGluProGluTyr 2304
Db 6415 TCAAAGAGCAGGAGATTTTTCGTAAGTTGAATCTCTCGTGGAGGTGATCTGTGATG 6474
QY 2305 AsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyLysLeu 2324
Db 6475 GCAGCTGCTCAGCAGAGNACTGAGATCTTTAATCTGCCAGCAGTTACTACGTCAGGCTCA 6534
QY 2325 MetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAla 2344
Db 6535 GTTAGCTCTAGAGGCCATTCTTTGCTGATCCTGCCAGT---AATCTTGGGCTCGAAGAC 6591
QY 2345 IleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGlu----- 2359
Db 6592 ATATCAGGAAGGCTCTCATGGGAAGCTTTGATGACAAAGTTGAGGATCATGGAGTTGCTC 6651
QY 2360 ---SerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuPro 2378

Db 6652 ATGTCCAGCCTATGGAGTA-----GTGCCTGGTACTGCCAACACCTCAGTT--- 6699
QY 2379 AlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGly 2398
Db 6700 -----GTGACCAGTGTGTGAGACACGAGAGAGGAAGAGGGGACCCCATCACCTCAT 6747
QY 2399 GlyGly---GlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerPro 2417
Db 6748 TCAGGAGTTTGCACCAAGCTGATCAGCAAGTCAACAGCAGGAGAAATCTTAAGTCTCCT 6807
QY 2418 AlaPro-----GlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSer 2435
Db 6808 ATACCTGGGCGAGGCTACTTAGGAACGGAACGGCCCTCTTCACTCTCTCTGTACATTCA 6867
QY 2436 GluGlyAspCysAsnArgThrProLeuThrAsnArgValTrpGluAspArgProSer 2455
Db 6868 GAAGGGGATTTACATAGGAGCAGGCCA-----GGGTGGCCCTGGGAGAGCAGGCCCTCT 6921
QY 2456 SerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyVal 2475
Db 6922 TCACAGGCTCAACTCAGTTTCTTATACCTCTGACTATGCGGATG-----CTC 6972
QY 2476 MetAlaSerProProProProGlyLeuProAlaGlySerGly---ProLeuAlaGlyPro 2494
Db 6973 AGCAGTACTCCACCAACACGATTCATGTGCTCCTCTCGGTGAACCAAGCAGCTCCT 7032
QY 2495 His-----HisAlaTrpAspGluGluProLysProLeuLysCysSerGlnTyrGlu 2511
Db 7033 CACCAACAGAACAGGATCTGGGAGCGAGAGCCTGCCCTCTCAGCACAGTACGAG 7092
QY 2512 ThrLeuSerAspSerGlu 2517
Db 7093 ACCCTGTGGATAGTGTAT 7110

RESULT 13
AF495886
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Xenopus laevis (African clawed frog)
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 8959)
Sachs, L.M., Jones, P.L., Havis, E., Rouse, N., Demeneix, B.A. and
Shi, Y.B.
Nuclear Receptor Corepressor Recruitment by Unliganded Thyroid
Hormone Receptor in Gene Repression during Xenopus laevis
Development
Mol. Cell. Biol. 22 (24), 8527-8538 (2002)
12446772
2 (bases 1 to 8959)
Sachs, L.M., Jones, P.L. and Shi, Y.-B.
Direct Submission
Submitted (27-MAR-2002) Lab. Physiology, UMR 8572 CNRS, Museum
National d'Histoire Naturelle, 7 rue Cuvier, Paris 75231, France
National Location/Qualifiers
1. 8959
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
499..7995
/codon_start=1
/product="nuclear receptor corepressor"
/protein_id="AA012658.1"
/db_xref="GI:20147787"
/translation="MSSSGYPPNQAFSTEQGRYSSHPVQYTFPSRHHQOQYSEYR
TAHLEASLIQQOQLRRRPSLSSEFPVSDRPPDRRSYDQYQTHISQNEHEALESKR

Qy	468	LeuTyrTyrTyrLeuThrIysLysAsnGluAsnTyrIysSerLeuValArgArgSerTyr	487
Db	1900	TTATATTATTATTAACTAAGAAAAATTGAAATTCCTTGTGAGAAAGAAATAT	1959
Qy	488	---ArgArgArgGlyIysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	506
Db	1960	CCTAAGAGGAGAGAGCGCAACACAGCAACAGATACACGCCCTTCCCAAGAAGAAAAAGGAA	2019
Qy	507	GlnGlnGlnGlnProMetProArgSerSerGlnGluGluIysAspGluIysGluIysGlu	526
Db	2020	ATTGAAAAAGTAGAGAGGAGAAAGCAGATAGAAATGACAAAAAGAGAGATGAAGAGGAGG	2079
Qy	527	LysGluAlaGluIysGluGluIysProGluValGluAsnAspIysGluAspLeuLeu	546
Db	2080	GAAGAAGAAAGAAAAAGAGGAGAAAGAGGAGCTTTAGGAGGGCGCCAAAGACAAAATTT	2136
Qy	547	LysGluIysThrAspAspThrSerGlyGluAspAsnAspGluIysGluAlaValAlaSer	566
Db	2137	-----GATGCTGTAGCTGAAGATGGGGATGAAAAAGATCAATCCATGCCCA	2181
Qy	567	LysGlyArgIysThrAlaAsnSerGlnGlyArgArgIysGlyValArgIleThrArgSerMet	586
Db	2182	AGGGTGCTTAAACTGCCAATATACCAAGGCAGACGAAAGGTCGATTACCATCCATG	2241
Qy	587	AlaAsnGluAlaAsnSerGluAlaIleThrProGlnGlnSerAla	602
Db	2242	GCAAGTGAAGCTGCTGGCGCAATGCTGTATTACACGGCAACTACTGCTCCAGTCACCCACC	2301
Qy	603	-----GluLeuAlaSerMetGluLeuAsnGluSerSerArgTfPThrGlu	617
Db	2302	ACTAGCACTGCCACCACCTGTTGCACCTGTTCTCTGTGGC-ACCACCTCCAGAGAAACCAAC	2360
Qy	618	GluGluMetGluThrAlaLys-LysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIle	637
Db	2361	TCCTCTCCACCACCAACAGCAAGAAAGTCTGGTGCATCATGTTCGNACTGGGGAGCAAT	2420
Qy	637	eAlaArgMetValGlySerIysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLys	657
Db	2421	TGCCAAGATGGTGGGATCAAAAAGCGAATCCCAAGTGTAAAAATTTCTACTTTAACTACAA	2480
Qy	657	LysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLysLeuIysMetGluIysGluAr	677
Db	2481	GAGACGCACAATCTAGACAATTTGCTTTCAGCAGCACAAAACAGAAGTCATCCCGGAGGCC	2540
Qy	677	gaAlaAlaArgIysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPhePr	697
Db	2541	TCGAGAGGACGCAGATGTCTCAATGTGACAGTATTGCTTCTACTGTATCTGCA-----	2595
Qy	697	oProValValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMe	717
Db	2596	-----CAGGAAGATGATGAAATGAAGCTCT-----AATCAGGAGGAAAAA	2636
Qy	717	tValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSe	737
Db	2637	CCCTGGAAGACAGTGAAGGT-----	2655
Qy	737	xGlyProAlaThrValAlaAsnAsnSerSerAspThrGluSerIleProSerProHisThrGl	757
Db	2656	-----CCAGAAAACAGTTTCAGACACTGAAGAAGTGCCTCTCTCTCTCTCTCTCTCTCT	2694
Qy	757	uAlaAlaLysAspThrGlyGlnAsnGlyProLysProAlaThrLeuGlyAlaAspGl	777
Db	2695	-----TCCCTGTCTGAAAGCAGCCAGATTAGCGATGATC-----	2727
Qy	777	yProProGlyProProThrProProArgArgThrSerArgAlaProIleGluProTh	797
Db	2728	-----TCTGTGTGATAGAACCACTTCTTCGGTATCCCATGGGAAGCTCC	2768
Qy	797	rProAlaSerGluAlaThrGlyAlaProThrProProAlaProProSerProSerAl	817
Db	2769	TCCAGAACAGATGCTGTAGCAAAACCTGCT-----TCAGATTTCATC	2810
Qy	817	aProProValValProLysGluLysGluGluGluThrAlaAlaProProVa	837

Db		: :	2811	TCCCTTCCTTATAGTA-----	2826
Qy	lGluGluGlyGluGluGlnLysProAlaAaGlu----	GluLeuAlaValAspth	855		
Db		:	2827	-----GAAATAACAAGCCCCAGAGACTCAATATGCAGAGCTTAAAGTG-----	2871
Qy	rGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGlyProAl	875			
Db	----AAGGAAGAATAAAGCACACAGAGAATGACGACAATGGAGGTGGAAGAGAGAAGTCA	2927			
Qy	aLysGlyLys-----		878		
Db	gagTGCAGAACTAAAATCAGACTTAACTCCAAAGTAGACACTAAAGCAGAACCATATGA	2987			
Qy	----AspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysG	897			
Db	GACAGAGCAAACCAACTAGTACAGAGTACTACACAAGTGAAACAGAGGTTTGACTCCAAAGA	3047			
Qy	uGlyGlySer-----GlyArgAlaThrThrAlaLysSerSerGlyAlaPr	912			
Db	GGATGACTCCATCGAAAAAGTTAAATGGATAGACAGAGCGCGCTGATATGGGATATGCACC	3107			
Qy	o-----	GlnAspSerAspSerSerAlaTh	920		
Db	ACCCCAAAACATTAGTCAGGCCAGCGCAAGAGTCACAGCCAGCAAGTACTCCAGCGCTAC	3167			
Qy	rCySerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSer--	939			
Db	CTGCAGTGTCTGATGAGGAGGTTCATGGAGAA---CCTGACAGGCCAAGGATGATTCTTT	3224			
Qy	ProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGl	958			
Db	CGGTTCAAGGCCTTCCCTGTGAATCAGCGGGA-----ACTTTTTTAAACA	3272			
Qy	nLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProlleGlnVa	978			
Db	AGGACCAATGGATTTACAACAGTTTACAGCACCAGCTGCTGTAACTCCAATG-----	3327			
Qy	lThrLysValHisGluProProArgGluAspAlaAlaProThrLysProAlaProAl	998			
Db	-----ACATCTGTAGTCTTGCAATTAACACGGGAAC	3362			
Qy	aProProProGlnAsnLeu-----	1005			
Db	AGCTATAGCAGCGACCACTTTCTATGTACCAGAGACATCTCTATGAAATAGTCTCCT	3422			
Qy	GlnProGluSerAspAlaProGlnProGlySerSerPr	1019			
Db	AGAGACAAAGACAGACAGAACAGCTGACTTTGGAAAAACAGATGTCAGCCAGCCC	3482			
Qy	oArgGlyLysSerArgSerPro-----AlaProPr	1029			
Db	AGNAAACATGCTTAAGAGCCCAATATGGATTGGGAAGGAAATACAGTCTATATGCCTTA	3542			
Qy	oAlaAsp-----LysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCy	1048			
Db	CACCTGAAGTCAAGCATCCTTTTGAGCATGAAGCACAATG-----CAAAATGT	3590			
Qy	sTrpThrSerClyLeuProPheProValProProArgGluValIleLysAlaSerProHi	1068			
Db	TGCTCGGTCTGTTTCACATACCGGCTCTCTCTAGAGATGTCGGCAGAGTTTCTCCACA	3650			
Qy	sAlaPro-----AspProSerAlaPheSeryrAlaProProGlyHisProLeuPro-----	1085			
Db	AGTAGATATGNAATCCGCTCGTTACTGTGTCCTCCAGTATTACAGCTGTCTCCACATCA	3710			
Qy	-----LeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSe	1103			
Db	AGTCATTACCTCTCTATCAGATAGCAAGGCTGCCAGTACTCGCCA-----ACCCG	3764			
Qy	rAsnProProLeuIleSerSerAlaLys-----HisProSe	1116			


```

Db 7780 -----GCCTGGGAAGACCGACCTTCTTCCACAGGTTCAACTCAGTTCCATCAATCC 7832
Qy 2466 oLeuileMetArgLeuGlnAlaGlyValMetAlaSerProProProGly----- 2483
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7833 TCTAACCATG-----GGATGCTGAACAGCAGCAGCCGCTCTATGTCTCTG 7880
Qy 2484 -LeuProAlaGlySerGlyProLeuAlaGlyProHieHisAla-----TrpAspGluGl 2501
    :|||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7881 CATTCACCATCCATGACTCAGACCTCTGCACACAGCAGCAGCCGAATCTGGAGAGGGA 7940
Qy 2501 uProLysProLeuLeuCySerGlnTyrGluThrLeuSerAspSerGlu 2517
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7941 ACCTGCACCTCTTCTTCTGAACAGTAGTAACTCTGTGCAGACAGCGAT 7989

RESULT 14
AC139377/c
LOCUS
DEFINITION Mus musculus chromosome UNK clone RP24-291G12, WORKING DRAFT
AC139377
VERSION AC139377.3 GI:38229465
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 194840)
AUTHORS Wilton,R.K.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 194840)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (01-FEB-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 194840)
AUTHORS Wilton,R.K.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Nov 8, 2003 this sequence version replaced gi:28209800.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M BB0291G12
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Assembly: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 191297 bases at least Q40
Consensus quality: 191797 bases at least Q30
Consensus quality: 192167 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1658: contig of 1658 bp in length
* 1659 1758: gap of unknown length
* 1759 3721: contig of 1963 bp in length
* 3722 3821: gap of unknown length
* 3822 192947: contig of 189126 bp in length

```

```

* 192948 193047: gap of unknown length
* 193048 193916: contig of 869 bp in length
* 193917 194016: gap of unknown length
* 194017 194840: contig of 824 bp in length.

FEATURES
         Location/Qualifiers
             1..194840
             /organism="Mus musculus"
             /mol_type="genomic DNA"
             /db_xref="taxon:10090"
             /chromosome="UNK"
             /clone="RP24-291G12"
             /note="assembly_name:Contig6"
misc_feature       1..1658
                    /note="assembly_name:Contig7"
misc_feature       1759..3721
                    /note="assembly_name:Contig7"
misc_feature       3822..192947
                    /note="assembly_name:Contig8"
misc_feature       193048..193916
                    /note="assembly_name:Contig1"
misc_feature       194017..194840
                    /note="assembly_name:Contig2"

ORIGIN
Alignment Scores:
Pred. No.:      1.32e-66      Length:      194840
Score:          3736.50      Matches:     1407
Percent Similarity: 20.21%    Conservative:  84
Best Local Similarity: 19.08%  Mismatches:    235
Query Match:      28.27%      Indels:      560
DB:               2          Gaps:         57

US-09-522-753-5 (1-2517) x AC139377 (1-194840)
Qy 694 AlaAlaPheProValValGluAspGluMetGluAlaSerGlyValSerGlyValSer 713
Db 136140 GCTGGGGCTCCAGCTTCTTCTGGAGGACAGCGCTAGGACTGATGAGGTGATCCGGAAAC 136081
Qy 714 GluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArg 733
Db 136080 CCTAGCTGTGCAGTT-----TCTGTTCGAAGG 136054
Qy 734 GlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSer 753
Db 136053 -----TCAGGGCCAGTA-----TCATTACGGGTGTGATTCCCCAG 136018
Qy 754 ProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeu 773
Db 136017 CCAGGGCTT---GCAGCTCAGAACAGATGCGCAGTAAGCCAGAGCTTGTGA--AGCCTC 135963
Qy 774 GlyAlaAspGlyProProGlyProProThrProProArgArgThrSerArgAlaPro 793
Db 135962 AGTGAGTGTGGCTGTCTCTGGGTGTTCACACCT-----GAAAGTCACAGGGGACCT 135909
Qy 794 IleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProProAla----- 811
Db 135908 ---GAGCCTTGTGGCCACAGCATGGGCAGCGCATGGGCTGGAGCCAGCTGCTGGGTT 135852
Qy 812 -----ProProSerProSerAlaProProProValProLysGluGluGluGlu 828
Db 135851 CTCAGGTCTCCCTGCACTCAGCAGCTGGGGCTCTCAGTCTTCCATCTGG-AGGGTGAAG 135793
Qy 829 GluGluThrAlaAlaAla-----ProPro 836
Db 135792 GAACAGACTGGCTGTGACATCCGAACAGTGCTACCTTTTGTCTGTTTGTGGCCA--- 135736
Qy 837 ValGluGluGlyGluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThr--- 855
Db 135735 -----CATGGGCTCCAATCCAGACATGGCCATATTTGGCAGGTACCCAGACACCTGA 135682
Qy 856 -----GlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGluGly 873
Db 135681 TCAACGGGGTCTGGTCACTGTCCT-----CAGTGTACCTTACAGGCTCATCAAGGC 135631

```

```
Qy 874 proAlaLysGlyLysAspAlaGluAlaGluAlaThrAlaGluGlyAlaLeuLysAla 893
Db 135630 CCGGCT-----GGAGCTCAGCCTACT 135610
Qy 894 GluLysLysGluGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGln 913
Db 135609 GACCAC-----TATAGTAGATG-GTTAGCGCACAGACTGCCCTCTCCCATGC 135563
Qy 914 AspSerAspSerAla-----ThrCysSerAlaAspGluVal 926
Db 135562 TCATCTAGATGACTGTATACACATCCATTTGGCAGACGTGCAGC----- 135515
Qy 927 AspGluAlaGluGlyLysAspLysAsnArgLeuLeuSerProArgProSerLeuLeuThr 946
Db 135514 -----CAGGAGGGCAGCAA-----GCCTGACCCAGCACTCTTGGT 135479
Qy 947 ProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeu 966
Db 135478 CCT-----ATACATCCAGTTGGTGGCATTTCTGGGGGAACCGTG 135440
Qy 967 LysGlnArgAlaAlaAlaLeuProProIleGlnValThrLysValHisGluProProArg 986
Db 135439 AAGCAAGATCAGCCCTC----- 135422
Qy 987 GluAspAlaAlaProThrLysProAlaProAlaProProProGlnAsnLeuGlu 1006
Db 135421 TGACACATGGCCCTGTGGTCCCTGTCTGGTCTGGCCATGAGCCAGTCCACATACCA 135362
Qy 1006 nProGluSer-----AspAlaProGlnGlnProGly-- 1016
Db 135361 AGCAGTAGCATGGGAACGTAGCAGCACCTCCACTGCTGCCAGCACCCCTCTTG 135302
Qy 1017 -----SerSerPro-- 1019
Db 135301 TCAGGGTCTGTTCACAGCTTGCAGTGTGTTTGAGGGTATGGAGATGGCAGCCCTCC 135242
Qy 1020 -----ArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaA 1037
Db 135241 CTGACTGCCACCTTAATCTTTTCAGCAGAGAAACCCCATCTTTCGGGTTTCCCAAC 135182
Qy 1037 agluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProVa 1057
Db 135181 TGAGGGCCCAAGACTACCGACTGAGCCGCCAGCTGGTCACTCGGCGCTGCCCTTCCCAT 135122
Qy 1057 lProProArgGluValLysAlaSerProHisAlaProAspProSerAlaPheSerTy 1077
Db 135121 CCCTCCACGGAGGTGATCAAGACTTCCCCACACCGCGTGACCCCTCTGCTTCTCTTA 135062
Qy 1077 xAlaProPro----- 1080
Db 135061 CACACCCCGGTGAGTGGGAGCGCACCAAGTTAGTCTGTGCACTCGGTGGGAGCCAA 135002
Qy 1080 ----- 1080
Db 135001 CTCTTTTCAGTGTGCGAGGGCTGCATGGTCCGATCGGATGAGACCTTGGCTGAAGGTGTG 134942
Qy 1080 ----- 1080
Db 134941 ACCTGTCACTCCTGTGAAGCTGGGCACCTTCTGAGACTCGGGAAAGCTTCCGCACTACC 134882
Qy 1080 ----- 1080
Db 134881 TGCACGCCCATAGGCAGCCCTTGGAGGAACTTGAGCTGGAAAGGGGAGGCGCAGCTGGGA 134822
Qy 1080 ----- 1080
Db 134821 GGTATTTTCCACAGTGACTAAGAGGATGGTCAATTAGAGGGAGCAGTGGCGAGTCTTGAT 134762
Qy 1080 ----- 1080
Db 134761 TGACAGATAGTGTAGAGACTGTTTGGGGGAGACCCTAGGTTTCTATTCTTGATGG 134702
Qy 1080 ----- 1080
Db 134701 TCCTTACCACCACCTAAAGTAATCCAGCGGGSCAGTGATTAACAGCATCAGCAGTGG 134642
Qy 1080 ----- 1080
Db 134641 CTGGCCAGGTAGAGCTGACCCATGGGACAAGTGTACTGAGGGAGCTTTGGAGAAACAGGA 134582
Qy 1080 ----- 1080
Db 134581 AAGGCATCCAGGAGGAGAAAGAACAGCATAGCATGTGCNAAGGGCCAGNAGCACTCTCGAG 134522
Qy 1080 ----- 1080
Db 134521 TCGGACACTTTGAGSCACCCGAGCATGTGTACAGCATAACTTTCCACCAGCAAGCC 134462
Qy 1080 ----- 1080
Db 134461 CAGGGATCCCACTGTCATGTCTGGAATCCAGCCCGAGGTGACTCTCTCTGAAGCAT 134402
Qy 1080 ----- 1080
Db 134401 AGGAGGGATAGCGTGTGCTGACTCTGGCAAGGTCCCTTTGGCCCTCTGAGGAGAAACAG 134342
Qy 1080 ----- 1080
Db 134341 GATCTAGGAGGGAGTGGAAATCGGGGTCTGTGGCTTAGAGCTTGGGTGTGATCGGGG 134282
Qy 1080 ----- 1080
Db 134281 TGGGGTGTGTGAGCGGTGCAGCTTTCAGAGATCCAGGACTGTGGGTTCCTGGTGTGCT 134222
Qy 1080 ----- 1080
Db 134221 CAAGCCACTGAAGCAGAGACATGTGGAGAGCGCTCCAAGAGACACTTCATATTTCTCTGC 134162
Qy 1081 -GlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProPr 1100
Db 134161 AGGTACCCCGCTGCTCTGGGCTCCACGATAGTGCCTGGCCCGCTCTGCCAGCTCCGCC 134102
Qy 1100 oThrIleSerAsnProProLeuIleSerSerAlaLysHisProSerValLeuGluAr 1120
Db 134101 C---ATCTTAACCCCCACCCCTCATCTCTCTGCAAGCATCCCGGCTACTTTGAGAG 134045
Qy 1120 gGlnIleGlyAlaIleSer----- 1126
Db 134044 GCAGCTGGGTGCCATCTCCCAAGGTGAGTACCTCTGGGGCTGGGTCTCCCATCAGAGGA 133985
Qy 1126 ----- 1126
Db 133984 GCCTTTTGAAGCCATGGGTGTTTCCAGGGATGGTACTGGGTGGAGCGGCTGGGCTCA 133925
Qy 1126 ----- 1126
Db 133924 TCCCTGCGAGTTGGGCGAGCTACGGGGCTACGGGTGATCTCAGGCTGTGATTTCCGCTGC 133865
Qy 1126 ----- 1126
Db 133864 CTCAGGCAGGTGACTAAACGTCTGGAGGCTCAGTTTCATCTCTGTTCACTGAGAGGGTG 133805
Qy 1126 ----- 1126
Db 133804 GGGAGCTGCAGCTGTACAGGCTATCGAGTGTACGTCATGAAGCCCTTGGTTCACCTTGC 133745
Qy 1126 ----- 1126
Db 133744 ATCCTACTGCTGCCCTCAGCACACCGGATGGGAAGCAGTCACTGTGTCTTTAGTGGGC 133685
Qy 1126 ----- 1126
Db 133684 ACTGGCCTTGGCAGTGTGTTGGCCGTCTCCATCTCCCTCAGCTGAGTGGGCGCAGACGCC 133625
Qy 1126 ----- 1126
```

Db	133624	CATCTGACCTGCATATCTTAAAGTCAGAGTGGGGCTCTCCCTGTGTCTGAAATCTGGT	133565
Qy	1127	-----GlnGlyMetSerValGlnLeuHisValProTyrSerG	1139
Db	133564	TGTGATCTGTTTCTTACCTTAGCAGGGGATGTGAGTCCAGCTTCGTGTGCTCAGTCAG	133505
Qy	1139	luHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProL	1159
Db	133504	AGCATGCCAAGGCCCCATGGGCCCTCTCACATGGGGCTCCCTCTCCCGTGGACCCCTA	133445
Qy	1159	ysLysLeu-----	1161
Db	133444	AGAACT--GGGTAGTTCCCAACACTTTGGTTCCTTACAACCTTGGGCTAAGGGGTGGGC	133386
Qy	1162	-----Alap	1163
Db	133385	AGTCCCTGGTGGAGGGCAGCTTCAGACCTCTGTGTATCTCTCACCCGTCCCTAGGCC	133326
Qy	1163	roPheSerGlyValLysGlnLeuSerProArgGlyGlnAlaGlyProProGluS	1183
Db	133325	CCTTCAGTGGAGTGAAGCAGGAACAGTTGTCCCTCGGGGTGAGCTGGGCCACTGAAA	133266
Qy	1183	erLeuGlyValProThrAlaGlnGluAlaSerValLeuArg-----	1196
Db	133265	GTCTGGGAGTGCCCACTGCTCAGGAGACCTCTGTGCTGAG--AGGTAAGGCTGGCACAGTC	133207
Qy	1196	-----	1196
Db	133206	ACCTTCACTCTAGAGAGCCTGAAAGCCCACTTCACAGCTGTTTCTCATCTTGAGAC	133147
Qy	1196	-----	1196
Db	133146	ACCAAGAAGACAGTTGGGTGGCTTTCAGACATGGGGTCTACGAGCTGGGTCTTTTGTG	133087
Qy	1197	-----GlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerT	1215
Db	133086	TTCTAGGAGACAGCACTGGGTTCGCCACCACTGAGGAAGCATCACCAGGGCTCCCCAGTA	133027
Qy	1215	hrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHis-----	1231
Db	133026	CCCCGGTGCAGACGGCCCC-----AGCTACAGAGGCTCTATCACCACAGTAGGTGCTG	132973
Qy	1231	-----	1231
Db	132972	GGGGCAGGTGTTACGGGCAGACATGCCAAGTGGCTTGGGATGGATGAGGAGACCTCCT	132913
Qy	1231	-----	1231
Db	132912	GGGAGACTAGAGACAATGGCTGAATGTTGATCACTGTGCCGTGTGGGGTCAGGCCTAGG	132853
Qy	1231	-----	1231
Db	132852	GTGATAAAGAGAACAAAGCTCCTGTGTCTTGGCAGTTAGTGTGGCAGACCAACCAC	132793
Qy	1231	-----	1231
Db	132792	AGATAGGATAGCCAGGCTCATACCAGGCTGTTCAGGAGCCACAGGCGATAGATGTCTG	132733
Qy	1231	-----	1231
Db	132732	GACTGAGAAAGTTACTGGGCCAGGAAGAGAAAGTGTGACATGTTATTATGGCCGAAG	132673
Qy	1231	-----	1231
Db	132672	AAATAGAGGAGCAGGGGTGGCAGCTGGCTTGACATATCACCAGTCAAGTAGGATTAAGC	132613
Qy	1231	-----	1231
Db	132612	CAGTGGTGGTGCAGACTTAGTACCACCGAGGAGGAAGATGGCGAGTTCCAGGCGA	132553
Qy	1231	-----	1231
Db	132552	ACCTGAACCTGTAGCTGAGAGACCTGTTTCTATGAACACAGTATGGTAGTATTCTCGATA	132493
Qy	1231	-----	1231
Db	132492	GCATCTCTGGGAGTCTCAAGTGAAGCAGGGGTGGCAGCAGCATTTGGCACTCGGAA	132433
Qy	1231	-----	1231
Db	132432	GCCAGGATGAGAGACATTGACAACAATCGGAGGGGAGTTACACATGTGTGGCTGTGCTC	132373
Qy	1231	-----	1231
Db	132372	TGTGCATCCTTGGGGGTGGGGAGTGGGTGTGAGGAGTCAACCCACAGATCCGG	132313
Qy	1232	-----GlyThrProAlaAspValLeuTyrLysGlyThrIle	1243
Db	132312	GTGCTGCCACTCCACCCCTACAGGGCACGCCCGCAGACGCTCTCTACAAGGGTACCATC	132253
Qy	1244	ThrArgIlelleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeu	1263
Db	132252	AGCAGGATCGTGGTGAAGCAGCCCAAGTGCCTTGACCGGGCAGCAGAGGACACCCCTG	132193
Qy	1264	ProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGlu-----	1281
Db	132192	CCCAAGGGCCATGTCTATAGGGCAAGAAAGCCACGCTCTATCTCTATGAAGGTGAG	132133
Qy	1281	-----	1281
Db	132132	GACAGAGCAGACAACACGAGGCTCTGGCAGTTAGGTGGGTGGCTATGCTGTCTCT	132073
Qy	1281	-----	1281
Db	132072	CCCTTTCTTCCGTTACAGGATGTCTTTTCTCAAGACAGACATAGCAAGCCACTGAGA	132013
Qy	1281	-----	1281
Db	132012	GCAACCACTTGTAGGCTGAGGCTTAGCAAGCAAGCACAATAGCAGATAGTACTTCCCTG	131953
Qy	1281	-----	1281
Db	131952	CCAGCTCGCCCTGCTGGGAGTGTGGTGAACCCCTTTGTTGATTAGGAGTCATGAAGAC	131893
Qy	1281	-----	1281
Db	131892	CGGACTCAACACACGCGCCACAGACAGCCGTTGTGGGTGTCTCCACTTTTCTGAGAGGG	131833
Qy	1281	-----	1281
Db	131832	GCCCTTGGAGTTGGCGTTGTAGCCTGGACAGTTGGGGTCTGGGCCCTAGGTTTCAATGCTC	131773
Qy	1281	-----	1281
Db	131772	AGGTCCTCTATGAGAAATACCTTCTCTGGCCCTGGTGAGACAAGGCAGCCCTTGTAGGG	131713
Qy	1281	-----	1281
Db	131712	CTGTAGGGTCTATAGAGGTCTCTTTAGAACAGGACAGAGCTGAAGTGATATCAGGGTGCA	131653
Qy	1281	-----	1281
Db	131652	GGCAGGGGAGTGACTTCTCTCTGACGGCATACAACCCAGATGAGGTGGGGCTGGGCTT	131593
Qy	1281	-----	1281
Db	131592	GGCGTTTCCCCCTACATAGGCTGGGAAAGCTGGGGCAGCCCTTGAGTGGCGTGACAGGG	131533
Qy	1281	-----	1281
Db	131532	AAGGGCTGTGGCAGAGCATGCCACCCTCTTCTCTGGCTAGAAAAAGTGAAGTGTGCTAC	131473
Qy	1281	-----	1281
Db	131472	CTAAAGTTGAGGAACCCGCCCTGGGAAGCATCCCCCATAAACAGGACCCCTCAACA	131413

QY 1281 ----- 1281
Db 131412 CTGGCTGTTTCTGGGACTGGGACTGCCAGTTCCTTATTCCTTTAGCCCCACAGCCCCCT 131353
QY 1281 ----- 1281
Db 131352 ATCTCTGAGTCAACAGTACCCCGTGTGTCCAGGGAACAGGACAGCTGCCTCAG 131293
QY 1281 ----- 1281
Db 131292 GTGCCAGCTGCATGCTCTGCTGTGACAGCCCCCAAGAGCTAGTAGCACTGAGTCAACAGCT 131233
QY 1281 ----- 1281
Db 131232 TGGAGCCTCCAGAGTCTTAGATTGTAACCTTCAGCCTGCTCTTTGTGGAGATGGGGTGT 131173
QY 1281 ----- 1281
Db 131172 CCTATGCGTCTTTTCTGGGCCCAGTGGGTCTGAGGCAGTTTCAGAGTGTCAACCCACCGTCA 131113
QY 1281 ----- 1281
Db 131112 GTCCAGCCCCAGCTGCCAGCCTTTCTTTACAAATATTTCTTTGTATTTCTCATATAGC 131053
QY 1281 ----- 1281
Db 131052 CAGCATGTTCCAGTCAACAGTGTATTAACAAAGCCACTCTGTCCCTGGCAGCGAGAGCAC 130993
QY 1281 ----- 1281
Db 130992 ACCTGTAACTGGCCTCACATGGTCCCTTGGACTCACACATATCCACITTTGTGCATGCAG 130933
QY 1281 ----- 1281
Db 130932 AGCTGAGGCGCTGTGGAGGCTTCCCTGCCGACTCAGACTGTGTGACATCTTTAGTGT 130873
QY 1281 ----- 1281
Db 130872 GCACATACCTATGGGTACAGAGCATACCCACCCTTACAAAGCCCGGTGGGCATG 130813
QY 1281 ----- 1281
Db 130812 CCCTGCTCTTCTGCCAATTGAGCTGGTGTGAGGGTCCCATCTGTGGGATGGCAGGCGCT 130753
QY 1281 ----- 1281
Db 130752 CTTTCCAGCTCTGCTATTATGGGGGCTTAGCACCGTGCCTGTAGTGCAGCCCCCTTGCAG 130693
QY 1282 ----- GlyMetSerValThrGlnCysSer 1290
Db 130692 CCGTCCACATCGTGTGCTTTCTCCGGCTCCAGGTGGTATGTCCGTGTACAGTGTCT 130633
QY 1291 LysGluAspGlyArgSerSerGlyProProHisGluThrAlaAlaProLysArgThr 1310
Db 130632 AAGAGGATGGAAGAGCAGCTCGGGCCACCCCATGAGACTGCGGCCCTTAAACGCACC 130573
QY 1311 TyrAspMetMetGluGlyArgValGlyValGlnIleSerAlaSerIleGlu 1328
Db 130572 TATGACATGATGGAGGCGGTGTAGGCAGGACTGTACCTCAGCCAGCATAGAGGTTAAG 130513
QY 1328 ----- 1328
Db 130512 TGTCCCGTGTAGCCACTTACCAGGACGAGGCTGCCAGGAACCTCCAGGACCCAG 130453
QY 1328 ----- 1328
Db 130452 GCCAGAAATGAAAAGAGGTGGCCTCTTGAACAGCTTGGAGGATCCAGTTAGTTGG 130393
QY 1328 ----- 1328
Db 130392 CTCATATTAACAGATAGAGAACTGAGGCACAACTTACTGAGAAATGGGGCCAGGAA 130333
QY 1328 ----- 1328

Db 130332 GTGGGGGTGCTTTCTTCTGCTGGTCTTGCATCATGGAATTTAGTGTAGACCAC 130273
QY 1329 ----- GlyLeuMetGlyArg 1333
Db 130272 AGCCCTCCAGACACCATATGTTGACACATGCTCTTCTCCCCAGGACTCATGGCCGC 130213
QY 1334 AlalIleProGluArgHisSerProHisLeuLysGluGlnHisIleArgGly 1353
Db 130212 GCCATC---CCTGAGCAGCACAGCCCC---CACCTCAAGGAGCAGCATCATCCGAGGC 130159
QY 1354 SerIleThrGlnGlyIle----- 1359
Db 130158 TCATCATCAACGAGGTAC-CTCCCTTATGTTGTGTGTTCCGCCCCCGGCCCAAGTACCAG 130100
QY 1359 ----- 1359
Db 130099 CTAAGGGGTGGAACACAGAGGGTGTGCTTCCCCACCCATGGTCTATAGGTGAGTGAATA 130040
QY 1359 ----- 1359
Db 130039 CTTGGCCCCGACTCCAAAGCTTTCAGCAAGCAGCCTTGTCTAGTGTGGAGGGGTCCA 129980
QY 1359 ----- 1359
Db 129979 GATTGAATGGAACTTTAGCTCGTGTCTCTGTGTGTAACCTGACAGTGTGTTTACTTAG 129920
QY 1359 ----- 1359
Db 129919 CATCTCAGCGCATATGAAGACAGCAGGCTTAGAAAAGTCAGTTCCGATGCCCGGGGCC 129860
QY 1359 ----- 1359
Db 129859 ACAGGAAGGAGGATTTGTGCTGACCTAGACCGAGCCAGGCAAGCAGTGCAGCAGCAGT 129800
QY 1360 ----- ProArgSerTyrValGluAlaGlnIleAspTyrLeuArg 1372
Db 129799 GCTGTGTGCTGTACAGGCAATCCCGAGTCTCTATGTGTGGAGCGCAGGAGACTACTTACC 129740
QY 1373 ArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProProProProProPro 1392
Db 129739 CGGAGGCCAAGCTCTTGAAGCGAGGAGGACACACACCCACCCACCCACCCACCCACCC 129680
QY 1393 LeuThrGluAlaTyrLysThrGln-----AlaLeuGlyProLeuLysLeuLysPro 1409
Db 129679 CTGACTGAGACCTTACAAGCCCGCCCGCCCTGAGCCCTCTGGGTCCCTGAGCTGAAGCCG 129620
QY 1410 AlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIlePro 1429
Db 129619 ACTCAGAGGGTGTGTAGCACTGTGAAGAGCGGGCGCTCTATCCATGAGATCCCG 129560
QY 1430 ArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGly 1449
Db 129559 AGAGAGGAGTGCGCCGACACACCTGAGCTACCCCTGCACACCGGCTCTTGAAGAGGGGT 129500
QY 1450 SerIleThr----- 1452
Db 129499 TCATCAC-CCAGGTACGAGAGTGCAGAGGAGGGGGCATGGGCAAGCCTGGGGGGTGTCT 129441
QY 1452 ----- 1452
Db 129440 ATGCTGCCCTCGTTGATTTGGGTATATAACTGAGGCCCCAGAGACATCAGCAGCTGGCCC 129381
QY 1452 ----- 1452
Db 129380 TGGTCACTGACAGTATTTGTGAACCAACCAAGTCTTATGTCCCTAACTTGTCACTGTG 129321
QY 1452 ----- 1452
Db 129320 TCCCCACGTCTCCACCTGCATGTATGTCAAGCCTTAGAGCCACTGTCTCTTAACCTTGGT 129261
QY 1452 ----- 1452

Db	129260	CATAGCTTGATGGGAAGAGAAAGACAAACAGCCAGACAAAGGGCGCTAGGAGGGGCTA	129201
Qy	1452	-----	1452
Db	129200	TCCCTGAGCCAGGACACACTCAGTGATGGCCAGAAATGATTCTAGAGAATTTCTTTGA	129141
Qy	1452	-----	1452
Db	129140	AAATTTAGTGCCACTCTTCTTTTGGGCCAGCACCTTAGAAGGCTAAGTCGGGAAGGTT	129081
Qy	1452	-----	1452
Db	129080	ATAAATTTGAGGCTAACCTAGGCTACAAAGCTGTACTACTAGCTTAATGCCAGTCACCACA	129021
Qy	1452	-----	1452
Db	129020	GTTATAAAGTTTAATAACAACACAGACATTGAGTAGCAATTGAGATCTAGGACCTAGCTG	128961
Qy	1452	-----	1452
Db	128960	TGCTCCGTGTCGCCGGCTTGGATCCCGCCCATCTCTCCATGATGTCCAATAATTCCC	128901
Qy	1452	-----	1452
Db	128900	TTCTCTGTGTGCGCTCTGAGACCTAACAGGGAATATGTAACTCCAGCTGGGACCCGAG	128841
Qy	1452	-----	1452
Db	128840	CCACAGCTGGCGTGGCAGCAAGTGCATCTGCTATGGNAATTCATAGCATTTGCAAGATGG	128781
Qy	1452	-----	1452
Db	128780	TGCAGTCCCACTGGGTGAGGAAGTTGTGTGACCTGAAGTTAGCCTCACTCTGGAACC	128721
Qy	1452	-----	1452
Db	128720	CCTGGGCGGTGGGTGACCTCGATGTGTAGCCAGACTCAAGTGAGCAGGGCGTCACGCC	128661
Qy	1452	-----	1452
Db	128660	TCTGGACATCATGTCTGCTAAAGGTCTCTGTCCATTGGCCCAAGGATTAGAGATGCC	128601
Qy	1452	-----	1452
Db	128600	CATCCTGTGCAGACATACAGGAAGAGGGCATCTGGCCAGCAAAATACAGACCTGCACCC	128541
Qy	1453	-----GlnGlyThrProLeuLysThrGlyAlaSerThrThrGlySerLysLys	1470
Db	128540	TCTCCCGAGGGCACCCCACTCAAGTACGACTCTGGGGCACCCCTCCACTGGCACCAGAAA	128481
Qy	1471	HisAspValArgSerLeuLeuGlySerProGlyArgThrPheProProValHisProLeu	1490
Db	128480	CACGAGTGGCTCATCATCGGACGCCCCGGCGCTTCCCTGCGCTGCACCCCGCTG	128421
Qy	1491	AspValMetAlaAspAlaArgAlaLeuGluAlaCysTyrGluGluSerLeuLysSer	1510
Db	128420	GACATAATGGCTGACGCGCGCACTGGAGCGTGCCTGCTATGAAGAGAGTCTGAAGAGC	128361
Qy	1511	ArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleVal	1530
Db	128360	CGGTGAGGACACGAGTGGTGGAGGGGCTCCCATCACAGTGGGGGTCCAGTCTGCTGTG	128301
Qy	1531	ProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPhe	1550
Db	128300	CCTGAACTGGGCAAGCCACGCAAGCCACTGACTTACGAAAGACACCGGGGACCCCTTC	128241
Qy	1551	AlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGln	1570
Db	128240	ACCAGTCACTGCGCAGTGGCTCCCTGTGTGACCAGGAGGACCCACGCCAGCGCTTCAG	128181
Qy	1571	Glu-----	1571
Db	128180	GAAGGTCAAGTGGGTGAGACTCCCTCAGGCTGCCAGCTGTGGGGAGGGGACAGCTCACTTC	128121
Qy	1571	-----	1571
Db	128120	CTACAGGATACCTGGCCACTGCTCTTGGCCAGCAGTCTGTCTCTGCTCCCTGCTCT	128061
Qy	1571	-----	1571
Db	128060	ACACTGCATACAGGGCCCCCTGAAAAAGTGTTTTTTAAATAAATCAAGGCTGAGGAGAAG	128001
Qy	1571	-----	1571
Db	128000	GCTCAGTAGTGCNTAAGCATGAAGACAGAGTTCAAAATCCCCAGCAGCCACATAAGCCA	127941
Qy	1571	-----	1571
Db	127940	GGCACACTCATGCAGTGTGTAATTTCTATGCTTTCACAGTGAGAGGAGGAGGACAG	127881
Qy	1571	-----	1571
Db	127880	GAGAGTCTCTAAAGGTCTCGGGCCATAAGTGGGAGTTAAATATCCAGACATGTTGGAG	127821
Qy	1571	-----	1571
Db	127820	AGTGAAGAACTGACCTTCAAAATTTGTCTTGACTTCTACTACACACCATGGCACATTTCTC	127761
Qy	1571	-----	1571
Db	127760	CATTTCATGTGCACACATACACAGAAAAATTAAGTAAATCGTTTGAATGAAGACAG	127701
Qy	1571	-----	1571
Db	127700	TTATCACTCCGTAGGCTTCATCTGTCAGTAGCACCTGCTGGTCTGAAGCCCTCGCTGTG	127641
Qy	1571	-----	1571
Db	127640	AGTCAACCCCGAGTTCTGTTCCTCTCTGATTTCTTTTACCTCCCTGACTTCCACTCT	127581
Qy	1571	-----	1571
Db	127580	CATTCTCCGATCTCTCTCAGCCTTTCCTGTGTGGCCCCCTCTTTAGAGGGCTCCACCT	127521
Qy	1571	-----	1571
Db	127520	TTGGTTCTGTTTCTAGCCAAACCCCTGTATATACCCCAACACCCCTTAATTCAGATGTCAGGG	127461
Qy	1571	-----	1571
Db	127460	CCACCAGGGAGGCTATGGCCAGTTCCAGCCACATCTCTGGCTAGGCTTTGTTAGGTGTC	127401
Qy	1571	-----	1571
Db	127400	AGATCTGTGTTTGATTTCATTGTCTCAGGGGGCACTTCCCAAGCAATTGAGATCATTTCT	127341
Qy	1571	-----	1571
Db	127340	CTCTCAAGCCTGTGTCTTCTCTACTTTGTGTTTCTATCTCCGACCTGGAGCAATGGCTCC	127281
Qy	1571	-----	1571
Db	127280	TTGGGGCTCAGGGTTCTTGCTGCTTTGTTTGGGGTGTCTCCAGCCCATGATAAAGTCA	127221
Qy	1571	-----	1571
Db	127220	GACTGTGGAGTGAATCCGGNAATGGTGACCCAGGGAGTAAATGGATATTTTCCAGGGGAC	127161
Qy	1571	-----	1571
Db	127160	CAACAGCAGAGGGGCAAGAGGCCCTGGGGTTTGTACAGCTCAGCCTGGTGGCCCCCACTG	127101
Qy	1572	-----GlySerLeuSerSerSerLysAlaSerCln	1581
Db	127100	AGGCTGAGGTACTGTGACCTGTCCCAAGGCGACCTCTCTATCCCAAGGCGTCCAG	127041

Qy 1582 AspArgLysLeuThrSerThrProArgGluAlaLysSerProHisSerThrValPro 1601
Db 127040 GACCGAGCTGACATCTACCCCGGGAGATGCCAAGTCCCAACACAGCACTGTGCC 126981
Qy 1602 GluHisProHisProHisSerProTyrGluHisLeuLeuArgGlyValSerGlyVal 1621
Db 126980 GAGCACCACCTCACCCCATCTCCCTATAGACACTTGTCTCCGGGGCGTGAAGTGTG 126921
Qy 1622 AspLeuTyrArgSerHisLeuProLeuAlaPheAspProThrSerLeuProArgGlyLeu 1641
Db 126920 GACCTGTACCGTGTGCATCCCATCTGAGCCCTTTGACCCCACTCCATACCCCGAGGATC 126861
Qy 1642 ProLeuAsp 1644
Db 126860 CCTCTGGAAGACAGTGGGTGTCTGGGCTGTAGATTGGGGCTTAATGATCTACTTCC 126801
Qy 1644 1644
Db 126800 TTGGGTGGGCTCTGGGGTCTGGAGAGTTGGGGTAGGCTTAGTATAGAGATCATCCCT 126741
Qy 1644 1644
Db 126740 GGTAGCCAGGAGGTTCAGGGGTGGGCTTGGGATCCTGTCTCTCAGTGGGCTTCT 126681
Qy 1644 1644
Db 126680 TTCCTGAGGATTCATCTGTACCTTGCAAGTGAAGTGGGCTCAAGTGAAGTGCATGACT 126621
Qy 1644 1644
Db 126620 TGAGCCCTTTTAAAGGCTCCTTTGTAGGTGTGAGGAGTATATTTTGACGATTATAT 126561
Qy 1644 1644
Db 126560 CTGTGAGCACTTCCAAAGCATTTGACAGACACACCCACATGATGAGGAGCAAGTCT 126501
Qy 1644 1644
Db 126500 AAGCTCTGTGTGTACTGGGGAGGATGTGGGTGACAGAGAGGAAGTGAAGTGGG 126441
Qy 1645 1647
Db 126440 GAGTATGAAGATAAGACATGGGCCACCTGACCTCCCTCCCTCCCGCAGCAGCGCA 126381
Qy 1648 AlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProPro 1667
Db 126380 GCCTACTACCTGCGCGCACTTGGCGCCCAAGCCCACTACCCACACCTGTACCCACCT 126321
Qy 1668 TyrLeuLeuArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleAsn 1687
Db 126320 TACCTCATCCCGGCTACCTGACACGGCGGCTCTGGAGAACCCCGCAGACCATCATCAAT 126261
Qy 1688 AspTyrIleThrSerGlnMetHisAsnThrAlaThrAlaMetAlaGlnArgAla 1707
Db 126260 GACTACATCACTCGCAGCAGATCACCACCAACGCTGCTCCGCGCATGGCCAGCGTGTCT 126201
Qy 1708 AspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGly 1727
Db 126200 GACATGCTGAGGGTGTGTACCGCGAGTCTCGTGGGCTCAATATTATGCGCGTGGC 126141
Qy 1728 ProArg 1729
Db 126140 CCAAGAGTGAGTTACCCAGGACTGCGCCCATCAGCAGAGCTTGCAAAAGACGAGTGC 126081
Qy 1729 1729
Db 126080 TTTGTGCACTGTGTAAGTGGGAAGAGGGGTACAGTTGCCAGGTAAGTGGT 126021
Qy 1729 1729
Db 126020 GCATGCTGGCTCTTGGGAGAGAGGTGTGCTTAGATCTTTTACGCTTAGCAGTGT 125961
Qy 1729 1729

Db 125960 GTAGTAGCAGGCACCTCTTGCTACCATCATGTTGTTCTGAGACTTTGGTATGAGCAT 125901
Qy 1729 1729
Db 125900 CTTTGTACATTCAGAGATCAGAACACAGGCTGGACCGGGTGTGTCATACCTTTTCTCTA 125841
Qy 1729 1729
Db 125840 AGTATCGGGAGGTTCAAAGCCAGCCTCAGCTACATAAAGCTTGAGGTGCCAGCCTGGGC 125781
Qy 1729 1729
Db 125780 TATGTGAGTCCATTTTAAATAACAATGTGTCTGTGTTTCCCGCCCGCCCAAAAAACAA 125721
Qy 1729 1729
Db 125720 AGAAAGAAATGAGTGTGAAGCCTTTTGGAGCACTGACTTCCAGAGCCTGTGAAGAGCCCA 125661
Qy 1729 1729
Db 125660 GGTCACTTTGTCAAGACTGTTGAGCAGCTCTGTGGGTTTCAGAGGGTGTGCGGTGTTT 125601
Qy 1729 1729
Db 125600 CGGTTCAAAGATGTTGGTTGAGAGACACTACCGGTGAGAGCCTTGAGAAGGAGCCTGG 125541
Qy 1729 1729
Db 125540 GGCAGCCTGGCCCTAGGAGGGGTACAGCAGGTGCAAGGCTCAGGCTACACCTTGG 125481
Qy 1730 1733
Db 125480 GGAACAGGAGGTGAGAACCTGCAGCTAATACCCCGCAGCCCTGCAGGCAATTATCGAC 125421
Qy 1734 LeuSerGlnValProHisLeuProValLeuValProThrProThrProGlyThrProAlaThr 1753
Db 125420 CTGTCCAAAGTGCACACCTGCGCGTGTGGTGCACCAACGCGCAGCCCTGCGCAC 125361
Qy 1754 AlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSer 1773
Db 125360 GCATCGACGCTGTGCTACCTCCCTCCCTGCGCCCGCCCGCCCTTTCAGAGCCCGCACGT 125301
Qy 1774 SerSerProLeuSerPro 1779
Db 125300 AGCTCACGCTGTCCCGCAGTGGCGCTGCGAAGTGGTGGTGGTTCGTGCTCACTCC 125241
Qy 1779 1779
Db 125240 CGATGCGCTGACAGTAAGTTCTGCGCCCTTGGCTCTGGCTCACACTCTGGCCCTGGGGTT 125181
Qy 1780 GlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluAr 1798
Db 125180 TCTAGAGGCGCCCACTACCTAGTAACTGACATCTTCATCGAGGCGGAGACG 125121
Qy 1798 GAspArgAspArgGluArgAspArgGluArgGluArgGluArgGluArgGluArgGluArg 1818
Db 125120 GGAAGTGGCGGGAACGAGAC 125079
Qy 1818 rThrThrValGluHisAlaProIleThrArgProGly 1830
Db 125078 CACTACAGTGGAGCATGCACCATCTGTGAGACCTGTGTAGGACACCCCGAGACCCCGCCCA 125019
Qy 1831 1836
Db 125018 CCGGAGTCTCATGGGCGACCTGGGCTGCGCCCTGACTCGGCGGCTGTG-TCACAGG 124960
Qy 1836 ySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerSerSerSerSerSerHi 1856
Db 124959 TACGAGCAGCAGCAGCGGGCTGGGGC 124908
Qy 1856 sAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnArgProSe 1876

Db	120545	TGGCCATTTTACAGGGAAATGTGCAGGGCCCGCTGTGTGGACAGTGGGGTCTCCTCATCATC	120486
Qy	2205	-----ArgSerProGluProAsnLys	2211
Db	120485	GGGTAGAACTTACCCCTCAGCTAGACGCTATCTCCATCAGGTCCCGCAGAACCCAGCAA	120426
Qy	2212	ThrSerValLeuGlyGlyGlyGlyLeuGluProValSerProGluGlyMet	2231
Db	120425	ACATCGGTCTCGGCAGCAGCGAGGATGCCATTGAGCCTGTGTCCCGCAGAGGGCATG	120366
Qy	2232	ThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGln	2251
Db	120365	ACTGAGCAGACATGCTCGAGCAGCTGCGTACCCACTGCTGTATCGAGACGGGAAACAG	120306
Qy	2252	ThrGlu-----	2253
Db	120305	GGCGAGCCAGGTACCATAGCACAAATCCCGCTTAGAGATGGTGGTGTCCATG	120246
Qy	2253	-----	2253
Db	120245	TACACCAAGGCTCGGCCCTCCCATATGGAGGCCATTAATCCCTCACAGCTCAGCGGC	120186
Qy	2253	-----	2253
Db	120185	TGATAGACCTTCAGGGGCTTCAGTTGGAGCACCTGTGTGTGGCATCTGTGGTCCTGGC	120126
Qy	2253	-----	2253
Db	120125	ATCCCTGTGTGGATCTCAGCCTGAGCCATGCCCTGTTTATCATCTTGTGTGTATCT	120066
Qy	2253	-----	2253
Db	120065	GGGAATGAACTAAGTTGAACACAGACTGGTGGGGTGGACATGACTTCAGGAGGGTGT	120006
Qy	2253	-----	2253
Db	120005	GACCTGCACCTGTCCATGTTCCCTGGGATTTGTGTGATCATGTAGCAATAAGGCATTGC	119946
Qy	2253	-----	2253
Db	119945	TGGGGCAATTACAGCGTACTCTGGGATGGGAACCACTGGTCTTGTAGCACCTCACGGTA	119886
Qy	2253	-----	2253
Db	119885	GCTCAAGTGTGAGTCCCTGTTGCTAAAGTCAACAATGCCAGCTGTGTGACCTGGTG	119826
Qy	2254	-----ProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAla	2270
Db	119825	TCTCCTGCTCCCTCGAGATGGGGCTCTAAGTCTCCAGGCAACACCCAGCCAGCCGAGCC	119766
Qy	2271	PhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluLe	2290
Db	119765	TTCTTCAGTAAGCTGACTAGAGCACTCCGCCATGTGTGAAGTCGAAGACGAGGATC	119706
Qy	2291	AsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyr	2304
Db	119705	AACAAGAAATCTAACACCCACACCGGAAACGAGCCAGATACAGTAAGGGGAGCCAGGG	119646
Qy	2304	-----	2304
Db	119645	CTAGAGCAGGACCGGAGGGAGATAGTACTCTTTAAACCTTACCATGGAGACCCCAACG	119586
Qy	2304	-----	2304
Db	119585	TTCTGCCAGTTGGTACTAACTGACAAACAGCATATAGCATTGTGGTAATGCCCTGCACAC	119526
Qy	2304	-----	2304
Db	119525	GTAGTGACACAGGAAAGCTGTAGAGACAGCCCTGTGTGTGGTGTCTGAGGAGCACAGCAC	119466
Qy	2304	-----	2304
Db	119465	ACTACTCCAGTTTCATGCTGCAGACCGCGTGCCTTTTACCACCGGCTGATGTGGGTGGACA	119406
Qy	2304	-----	2304
Db	119405	AGGACTCATGCTCTTAGCGCAGGGTCACATGGGCCATATGCTTACAAGAATGCACCGGAA	119346
Qy	2304	-----	2304
Db	119345	AGCAACACATGACAGCTTTGCACTGTCCCTCCCGAGGTGGCTGGACAGACAGGCAAG	119286
Qy	2304	-----	2304
Db	119285	GTTCCTCTCTTCTGGAGTATAGAAGCCCTCACAGGTAAACAGAGGCCTAAGGTAGTGACCA	119226
Qy	2304	-----	2304
Db	119225	ACCGCTGAGTAAACCTGGAGCATCAGGTGCTGTTGTGCCAGACATTCAACTCAGAC	119166
Qy	2304	-----	2304
Db	119165	TGTACGGGGACAAGCAGGGGTGCAGATCTCTTAGGACCTGTGTGAGCTCAGCCCTTCCCA	119106
Qy	2305	-----AsnIleSerGlnProGlyThrGluIlePheAsnMetProAla	2318
Db	119105	CCTTGTACCTGCTCTAGATATTTGCCAGCCTTGGGACGGAATCTTCAACATGCCCGCC	119046
Qy	2319	IleThrGlyThr-----	2322
Db	119045	ATCACTGGAGCAGGTAACCTCCTGTGTGCCATGGCAAACTCCCTTCCCGAGTGTGTG	118986
Qy	2322	-----	2322
Db	118985	ACTTGGGACAAGAGGCTTTAGCCTCTCTCCCTGTCTCAGTGGACCCAGAGAATGACACAGAT	118926
Qy	2322	-----	2322
Db	118925	GGGAGCCTCCACTGCCTAGGGTCTCCCTACATGAGGAGTGCCCAAGTTTGTAGGTGAG	118866
Qy	2322	-----	2322
Db	118865	GGTCCCTCAAGTGGTGTAGGTACACCCCGGAAACACTTTGGCCAGTTGTGGGGCTTGG	118806
Qy	2322	-----	2322
Db	118805	GGCTGGGGCAGCTCACCAAGTGTGACCTGAGGGATTGGCCTTGGGGTACAGAGATTGAG	118746
Qy	2322	-----	2322
Db	118745	GGTCCCGCAGGCATCTTTTACACTAAAGTCTCTGGCCATTTTCTAAATCAGATGCAGGCA	118686
Qy	2322	-----	2322
Db	118685	GTAATGTAGGGAAAGTGTCAAACTCACGTGCAGTTTGTGGCAGAAAGTCTCAGGCTTTC	118626
Qy	2322	-----	2322
Db	118625	AGTATGTGTCTGGGGCTTAGAGCATCCGCTTCCCGTGTATTATCACGTGTGCCCTCTG	118566
Qy	2322	-----	2322
Db	118565	GGAAAGAGCTTGTGTGGGAAACAAATGGGCTCGGTGACTGTGTGTCA CAGGGGAAGTGTG	118506
Qy	2322	-----	2322
Db	118505	GGGACTTAATCCAGGCCAGAGGAACAGCAAGGGTCCCTACTTCTTATTTGTGTGGAGC	118446
Qy	2322	-----	2322
Db	118445	CTTGGGCTTAGGGAATAAGGTGGCCAGATTAAACCTTGGGGACATTTCCAGAGACCAACA	118386
Qy	2322	-----	2322
Db	118385	GGACCTGTAAAGGGATGGGGTAGGATCTGTGGCAGGACAGCCTTGGGGCTTGTGTGAGC	118326

Db 116165 CACCAACATGGGGCTAGAGCCATTATTAGAAAGGCACTCATGGGTAAATATGATCAGTG 116106
Qy 2357 pGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLe 2377
Db 116105 GGAAGAGCCCGCGCTCGGCGCAATGCTTTTAAACCTCTGAATGCGAGCGCAGTCT 116046
Qy 2377 uPro---AlaAlaMetProLeuThrAlaAlaAspGlyArgSerAspHisThrLeuThrSe 2396
Db 116045 GCCCGTGTCTGCTATGCCCAACCACTGCTGACGAGCGAGTGACCAACGCACTCACCTC 115986
Qy 2396 rPro----- 2397
Db 115985 GCCAGGTCTGAGGCTGCCCGCTGCCCGCGCGCCGCCGCCGCCCGCTGTGCTCCTCTT 115926
Qy 2397 ----- 2397
Db 115925 ATCCCGGTGTGTTAATCTCTACCTTCCCGCCCTCAAGCTTGGGAGCCTGTGACCTTA 115866
Qy 2397 ----- 2397
Db 115865 TTTTGTGCGTTTGACATCTAGTTTGTCTAATCTAAAGCTGTCTGACACCCCATGTTGCG 115806
Qy 2397 ----- 2397
Db 115805 GAGGCGCCCCCTTTTACCCCCAGCACTGTCTGCTTTTCCCGTCTCAGATCACTCCCC 115746
Qy 2397 ----- 2397
Db 115745 ACATCTCCCTGCTGTGTGACAGGAGTACAGCGGCCAGGCCACCACCTACTGTGCCA 115686
Qy 2397 ----- 2397
Db 115685 GGGCTCTGGAGAGAGTCTAGTGTGGCGTGAACCCCTTGAAGGCCAGACAGCACTCC 115626
Qy 2397 ----- 2397
Db 115625 CCACCTTCCAAATGTCGTGAGCACTGTTGTTTCCAAAGCAGAGACACAAGAGGT 115566
Qy 2397 ----- 2397
Db 115565 GCCCTGTGTCCCTTCTAGGTCAGGCCAGCCCGCTGTGGCGCTCTCAGCAGAGCCT 115506
Qy 2397 ----- 2397
Db 115505 TCAGTCGGCATAGTGTGAGTGTCTGTGTCGCAATGCTGTATCCAGCAAGATCTGAC 115446
Qy 2397 ----- 2397
Db 115445 ATTCCTCAGTGAGGGGTACAGTCTCTGTGGAGCAGCCTAAGGGAATAGTCATCAATCTCT 115386
Qy 2397 ----- 2397
Db 115385 CTCAATGATGTTTCATGCTTCTTTCACACCAAGGGCTGGCTCTGGCCATCTGGTAGA 115326
Qy 2397 ----- 2397
Db 115325 CCTCTAGCTCCTGGCTCAGGGTTTCTCACACTAGGTTTCTGCTTTGGGGCCCTAGGA 115266
Qy 2397 ----- 2397
Db 115265 GGCAGGTGGTCCCAAGGCAGAGATCAGCCCCACTGGCCCTCTCAGTGTGGTGTGGGT 115206
Qy 2398 -----GlyGlyL 2400
Db 115205 GTCTACTGTCAACCCCGTCTGACTTGGCCCTTCCCACTTTGTCCCTGAGGTGGAGG 115146
Qy 2400 YGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGl 2420
Db 115145 TGGGAAGCAAGGCTCTCTGACAGCTAGCAGCGGAAAGCCAAAGTGCAGCACCAGG 115086
Qy 2420 YLeuAlaSerGlyAspArgProProSerValSerValHisSerGluGlyAspCysAs 2440
Db 115085 CTTAGCGTCCGGAGACCGACCCCTTCTGCTCTCAGTACACTCAGAGGGGGGACTGCAA 115026

2440 nArgA-gThrProLeuThrAsnArgValTTrpGluAspArgProSerSer----- 2456
115025 TCGCCGAACACCACTACCAACCGTGTGTGGAGGACCGGCCCTCATCTCGAGGTGGTA 114966
2456 ----- 2456
114965 TCAGTAGGAGTAGAGATGATCCAGGGTTGGGGGTACTCTGGGCTACCCAAAGAA 114906
2456 ----- 2456
114905 GGTCCCTGTACCACTTACCAAGGAGGATATAGGAAGAACTATCAGAGGGTCTGTG 114846
2456 ----- 2456
114845 TAGTGGCCAGACCTTCCAGGCATGGCTCGTCATACAGGGGAGGCTGGCAGCCTTCTTG 114786
2456 ----- 2456
114785 CCTGTACGCACTGTAGGATTTTGACAGGACAGTAAGGAGCAGTAGAGACAGGACAGG 114726
2456 ----- 2456
114725 GAGAGCGAAAGGTTTACTGGAGTCCAGTCAAACGCTGTAGGGCTGTGAGCACAGAG 114666
2456 ----- 2456
114665 GCAAAGGAGGAGCGCGTTTTCGAGAACTCCAGAGTTAGAGGTCTCTGGCCCAAGAGACC 114606
2456 ----- 2456
114605 CAGGCTCTGTTTCTTATATGGAAGAGCGGTCTCTAGGTGCCTACGTGTAGCGCTTGCA 114546
2457 -----AlaGlySerThrProPheProTyAsnPr 2466
114545 GAGCTCTCTCAGCACTCTCCCATCTTCTCGAGGGTCCAGGCATTCCTCTACACCC 114486
2466 oLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAl 2486
114485 TTTGATTATGAGCTACAGCAGGTGTCTATGGCTCTCCCGCCCGCCCTGCGCTTGGGC 114426
2486 aGlySerGlyProLeuAlaGlyProHisHisAlaTrpAspGluProLysProLeuLe 2506
114425 AGGCAGCGGGCCCTAGTCTGTCGCCACCAAGCTGGGATGAGGAGCCCAAGCCACTGCT 114366
2506 uCysSerGlnTyrGluThrLeuSerAspSerGlu 2517
114365 GTGTTACAGTATGAGACACTCTCGGACAGCGAG 114332

RESULT 15
AC097560/c 254449 bp DNA linear HTG 10-MAY-2003
LOCUS Rattus norvegicus clone CH230-157E10, WORKING DRAFT SEQUENCE.
DEFINITION AC097560
ACCESSION AC097560
VERSION AC097560.7 GI:30521202
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 254449)
AUTHORS Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsebrook, S., Amin, A., Angiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyne, M., Cree, A., D'Souza, U., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulvik, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Loulseg, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaekelemeh, O., Okunolu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Popper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scheter, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (19-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 254449)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:23265645. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Genome Center
Center: Baylor College of Medicine
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GIBP

Center clone name: CH230-157E10

----- Summary Statistics

Assembly program: Atlas 3.0;
Consensus quality: 245355 bases at least Q40
Consensus quality: 247474 bases at least Q30
Consensus quality: 248954 bases at least Q20
Estimated insert size: 254924; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* 1 254449: contig of 254449 bp in length.

FEATURES

Source

1. 254449
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-157E10"

misc_feature

3075..3548
/note="clone boundary
clone end:Sp6
site:EcoRI"

misc_feature

end sequence: BH283564"
complement (251005..251579)
/note="clone boundary
clone end:T7
site:EcoRI
end_sequence: BH283551"

ORIGIN

Alignment Scores:

Pred. No.:	2,59e-61	Length:	254449
Score:	3479.50	Matches:	1345
Percent Similarity:	18.72%	Conservative:	71
Best Local Similarity:	17.78%	Mismatches:	167
Query Match:	26.33%	Indels:	5989
DB:	2	Gaps:	43

US-09-522-753-5 (1-2517) x AC097560 (1-254449)

QY	905	ThrAlaLysSerSerglyAlaProGlnAspSerSerAlaThrCysSerAlaAsp	924
Db	234429	AGCACAGGACTGCGCTGCCACCCCTCATCTGGAGTAGTCCCGTCCGACCATCCCTT	234370
QY	925	GluValAspGluAlaGluGlyAspLysAsnArgLeuLeuSerProArgProSerLeu	944
Db	234369	GCCATAGACCACAGTCACGAGGTGACAA-----GCCTGACCAGCACT	234325
QY	945	LeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLys	964
Db	234324	CTTGGTCT-----GGACACCCCAATTGGTGACATTCGCTGGGAG	234286
QY	965	GlnLeuLysGlnArgAlaAla-----AlaIleProIleGlnValThr	979
Db	234285	ACCGTAGAAGCAAGATAGCCCTCTGACACATGCTGTGCTGCTGCTGCTGCTGCTG	234226
QY	980	LysValHisGluProProArg-----GluAspAlaAlaProThrLys	993
Db	234225	TCAGCCACAAGGCGAGCCCATACCAAGCAGTTATATGGAACATGACACACCTTGCTG	234166
QY	994	ProAlaProProAlaProProProProGlnAsnLeuGlnProGluSerAlaProGln	1013
Db	234165	ACATGTCCCAATGGCCCTCATGTCTAGGATACGTGTATAGCTGACCATGAGATGTCT	234106


```
QY 1014 GlnProGlySerSerProArgGlyIysSerArgSerProAlaProAlaAspLysGlu 1033
Db 234105 AGCCCTCCCTGACTGCCAACTT-AATCCTCTTTTCAGCAGAGAAAGCTTCATCTTTCC 234047
QY 1034 AlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProCysTyrThrSerGlyLeu 1053
Db 234046 GCCTTCCAGCTGAGGGCCGAGCTTCCTACTAGGCCCCACCGCTGGACATCGGGTCTG 233987
QY 1054 ProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspProSer 1073
Db 233986 CCTTTTCCCATTCCTCTCGGAAAGTGATCAAGACTTCCCCACACGCTGCAGACCCCTCT 233927
QY 1074 AlaPheSerTyrAlaProPro-----1080
Db 233926 GCCTTCTCTACACACCCCTCGTGGTGGGACCCACCACCCGATTAGGTCTCTGCAGTCAG 233867
QY 1080 -----1080
Db 233866 TGGCAGTCGGCTCTTGGCACTGGCGCAGGGCTGCATGGTGTGCTGGTGGGTGAGATGAGA 233807
QY 1080 -----1080
Db 233806 CTTTGCCTGGAGTTTGGCGAAGCTGGCACCTCTCTGAGAACATCTACACACCCCTGTA 233747
QY 1080 -----1080
Db 233746 GTGSCCTTAGGCAGCCATGGAGGAGAAAGACGACGAGGAGATACTTTCCACCGTGACT 233687
QY 1080 -----1080
Db 233686 AAGAGGACGTCGTTAGAGGAGGGACGGCAGTCCTTTGATAGTTTTCAGGAAGCCACTA 233627
QY 1080 -----1080
Db 233626 GGTTTTCTATTCTTGATGGTCTCCACCACGATAAAGTAAATCCAGCCAGGGGCCACCGG 233567
QY 1080 -----1080
Db 233566 CAATGGCAATGTCAGCAGTGGTGGCTTAGGTACAGCTGACCTGAGGGACATGTGCACT 233507
QY 1080 -----1080
Db 233506 GTTTGCTGGAGCTTGGAGAAACAGAGAAAGGCAATTCGCCGAGAGGAACAGCGTGTCAA 233447
QY 1080 -----1080
Db 233446 AGGGCCCAAAACACGAGTTGGGACACTTGGACACCAAGTCAAATCTGTCAAGAGA 233387
QY 1080 -----1080
Db 233386 AAATTACTACCCAGCCAGCCAGGGATCCCAGTCACTGCGCACTGAAATCCATGCCGA 233327
QY 1080 -----1080
Db 233326 GGTGACTCTCCTAAGCGGGGGGGGGGGGTATTGTAGGCTCACAGTTAGGGAGAGGT 233267
QY 1080 -----1080
Db 233266 TGGCCTACTGACTCTGGGAAAGATCCCTTCGGACCTTGGGAGGAGAACAGAACTAGGGA 233207
QY 1080 -----1080
Db 233206 GGGGGTAGAAGTGGGTGCTGGCTTAGAGGCCTGAGTGTGGATCAAGAGGTGATGTGCA 233147
QY 1080 -----1080
Db 233146 GCGGTGAGGATCCAGGACTGGGGTTCTTGTGGAGGCTCAGGTCAATTGAAGCAGAGGTA 233087
QY 1081 -----1086
Db 233086 TATAGGAGCCCCCACATATACACTTCTGTGTTCTTGGAGGTCAACCTCTGCGCTCT 233027
QY 1086 uGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnProPr 1106
Db 233026 GGGCCTCCAGATAGTCCCGCGGCTCTGCCACGCTCCCTCC-ATCTTAACCCCC 232970
QY 1106 oProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSe 1126
Db 232969 ACCCTCATCTCTCTGCCAAGCACCTCTGGTGTACTTGGAGGAGCAGCTGGGTACCATCTC 232910
QY 1126 r-----1126
Db 232909 CAGGTGAGTACCTTTGGGCTGGGTCTCCCTTTCACGGGAGCCCTTCTGAGCCGTGGGA 232850
QY 1126 -----1126
Db 232849 TGTGCCAGGGTGTGTCTGGCTGGACTCAGCCCTGCTGCTGTGGGTACGTGTGTGCACA 232790
QY 1126 -----1126
Db 232789 GGGCGTGTGTCCAGTGCCTTAGCAGGTGACTCACTGTCTGTAGGCTTCAGTTCATCCT 232730
QY 1126 -----1126
Db 232729 CTGTTCACTGGAGGTGGGAGCTGCAGCTGTGCAGGCTGTAGAGTGTCTGTCTGTC 232670
QY 1126 -----1126
Db 232669 GTAAGCCCTCGTCACTCTGTCATCTTACTGACAGTCACCAACTCACCATGTTTGCAC 232610
QY 1126 -----1126
Db 232609 AAGCAGTCACACTGTCTTTCTGTAGACTGGCCTTGGGCTGTGTGTGGCTGGCCACA 232550
QY 1126 -----1126
Db 232549 GTCTCCAGTCGAGGGCCGCGCCGCTCTCTCACTTGGCTATCCTTAAGTCTTAAGCA 232490
QY 1126 -----1126
Db 232489 AGGCTCTTCTGTCTGAATCTGGTCTTCCCGTGGCTGAGTCTGTGTGTATCTATCTC 232430
QY 1127 -----GlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaIleVal 1143
Db 232429 TTGCGCCTAGCAGGGAATGTCACTGTCAGTTCGCGTCCGCCACCTCAGAGATGCCAAGGC 232370
QY 1143 aProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLys-----1160
Db 232369 CCCCATGGGCCCTCTCACCATGGGACTGCTCTCGCTGTGAGCCCTTAAGAAGCTGGGTGA 232310
QY 1160 -----1160
Db 232309 GTGCCATGACCTTTTGGTTCTTATGCCCCGGGGGAGTCTTGGCCCTGGGAACCTAGAG 232250
QY 1161 -----LeuAlaProPheSerGlyV 1167
Db 232249 GGGGGCAGCCTGAGACCTCTGTATACCTCTCGCCACACCCCTAGCCCCCTTCAGTGGAG 232190
QY 1167 allysGlnGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValP 1187
Db 232189 TGAAGCAGGAGCAGTTGTCCCTCGGGCTCAGCTGGGCCACCTGAAAGTCTGGAGTGC 232130
QY 1187 roThrAlaGlnGluAlaSerValLeuArg-----1196
Db 232129 CTACTGCTCAGGAGACCTCTGTCTTAAGAGTAAGTCTGGCCCCCAGTCACCTTCAGTCTTA 232070
QY 1196 -----1196
Db 232069 GAGAGCCTGAAAGCCCACTTTCAGAGTTGTCTCTTACCTTGTACTCCACAGGACACGCT 232010
QY 1197 -----GlyThrAlaLeuG 1201
Db 232009 AGGTGGCTTTTACACTTGGTCTAAAGAGCTGTCTTATGCCCTTCTTAGGACAGCACTGG 231950
QY 1201 lySerValProGlyGlySerIleThrLysGlyIleProSerThrArgVal-----1217
```

Db	231949	GCTCCACTGCCAGTGAAGCATCACCAAGGGCTCCCGAGTACCCGGGACTGCAGATGGCC	231890
QY	1217	-----	1217
Db	231889	CCAGTACCGAGGCTCTATACCCACGTAGTGTCTGGGGCAGGGTGGGATGGGCAGGCA	231830
QY	1217	-----	1217
Db	231829	TCCTAGGAGGTTGGGATGGGATGAGAGGACCTCCGGGGAGACGAGAGATCTTGACCTGA	231770
QY	1217	-----	1217
Db	231769	ATGTTGGTCACTGCGGTGTGGGTGAGGCTAGGCTAGGTGATGAAGAAAAACAGGTGTCTATGT	231710
QY	1217	-----	1217
Db	231709	GTCTCCAGCAGTTGGGTGTGGCAGAGACCAACACAGAAAGGATAGCCAGGCCCTTAACC	231650
QY	1217	-----	1217
Db	231649	AGACTGTCAGGAGCCATAGATCATAAAGTGTCTGGACTGAGAAAGTTACCAGGTGACCCC	231590
QY	1217	-----	1217
Db	231589	AGTGACATGTTGCCATGGCCAGAAAGACAGAGGGGACAGAGGTGGCAGCTGTCTTGA	231530
QY	1217	-----	1217
Db	231529	CAAAATCACCAGTCCGTAGGCATGAGCGAGTGGGTACCGGCTTAGTACCCACGGTGT	231470
QY	1217	-----	1217
Db	231469	GTGGCCAGCATACAGCCAGCGAACTTCCAGTCACGGACACACCACAGAGGTCCTT	231410
QY	1217	-----	1217
Db	231409	TTGACTCACTGTAGGGGACAAATGTCCATTCATGCGCAGGCTTCCCTGAAGTCTCTGG	231350
QY	1217	-----	1217
Db	231349	AAATCTTGGGCTTTCAAAGGCAAGCTCAGAACTGGAGCGGAAATTCAGTCATGTCCT	231290
QY	1217	-----	1217
Db	231289	TCTGGCAGTCCAGTAGCCCTCCCTGACTCAGTTCCTTCCCTGTGACAGAAATGTTCTTCC	231230
QY	1217	-----	1217
Db	231229	AGGATAGGTACTCCAAGGAGCTCTTAGAGTGCCAGGTTTGCACACAGAGCTTCGCAGGGC	231170
QY	1217	-----	1217
Db	231169	TCAGGGATGCCGGCTCAGGCTGAGAACTCCCTTTCCCTGGTTAGGGCTTGTGGAGAT	231110
QY	1217	-----	1217
Db	231109	GAGGTACGCCCTCTGGGGCTATCTGTGTGGATGATGATGCCAGGCATTCATGTGACACG	231050
QY	1217	-----	1217
Db	231049	GCTAACAGGTGGAGACCACATCCCACTGTGTGCTGACCGCCTGGAAACCAAGGCCAC	230990
QY	1217	-----	1217
Db	230989	AGAGCCCCATGTCCCCTAGCTCCCAGGATTAACATCAGGCTTGGAGCGGCTCGGG	230930
QY	1217	-----	1217
Db	230929	TGGGCCCTGTGGCTCCCTGTATCTTTGGTTTGTCCAGTGAGTTATACGGAGCAGAAC	230870
QY	1217	-----	1217
Db	230869	CATGGCTGGCTGTATTTTCGGTTGTGACAGGCCCTCTAAGCTTGGAGACTCAAAACATC	230810
QY	1217	-----	1217
Db	230809	AATACCCCTTTTTCATCTGGAGCCTGCCTGACCTTTCTCACCGTATACCTGGCTAGCAA	230750
QY	1217	-----	1217
Db	230749	GCAGCTTACAGCCACTTGAGAAGGCATTTCCCGACACACCAGAAATCACCCCGCATGCC	230690
QY	1217	-----	1217
Db	230689	CCACGGGGACCAACCCACAGAAACAAGTCCACGGGAGACAAACATCACTGAATCTGTTCC	230630
QY	1217	-----	1217
Db	230629	TGAGGTCAATAATGAAGAGAGAACTCAGGATTTGGGGCCCTCTGAGAGCAAGTTAGAT	230570
QY	1217	-----	1217
Db	230569	CCTCATAGGATGGATACGTGCTGACTAATCTCAGGAAAGCAGGTGACTTGTATTAAAC	230510
QY	1217	-----	1217
Db	230509	AGCAGTTAGCAACTTGGGGTGATCCTTATATCCCCCATGGGCTGTCTCAGAAAGCTG	230450
QY	1217	-----	1217
Db	230449	TACTAATGGCCACACTCCGGTTCAAGAGGTCACTAAGTGGATACAATCCCAACAAACC	230390
QY	1217	-----	1217
Db	230389	GCATGCCGTTGGTTTGAGAACTCCTGGGGTGGGAGCTTACTCCCTCTAAATCAGTGTTT	230330
QY	1217	-----	1217
Db	230329	CTTAGTTGCCCGACCTGCCCCAGGACATCAGGGGCAATTAAGGGTGTGTTTGGTGT	230270
QY	1218	-----ProSerAspSerAla-----	1222
Db	230269	TATAGTGGACAGAGGTGGAAGGGCGTGAGTCACCTAGTCTATTCACAGACCAGGAAGAGTG	230210
QY	1222	-----	1222
Db	230209	CCTGGCTGCTCTACTACCTCAGTTTCCTGTGTTGGTGGCTCGGACAGCAGTCCAGC	230150
QY	1223	-----lleThrTyr-----	1225
Db	230149	TCTTGAAGGCTCAGGACGAAAGGCTGGTGAGTTCCAGGCTAACCTGAGCTATATTG	230090
QY	1225	-----	1225
Db	230089	CCAGACACCTTGTCTCTATGGAACAGTATTCCCGATAGCGTCTCTCGGAGGTACAAGTG	230030
QY	1225	-----	1225
Db	230029	AGAGACAGGGTGTGGCAGCAGTGTGACACTGGGAAGCCAGGATGAGTGAACGGTGGCAA	229970
QY	1226	-----ArgGlySerlleThrHis-----	1231
Db	229969	TGGTGGGGGGGAAGTTACACATGCATGCTGTGTATACCTTGTGGGTGGGGGAGTAGG	229910
QY	1232	-----GlyThrProAlaAspValIleuTyrL-----	1240
Db	229909	TGATGTGTCAGGAGTCAACCACTCTACCCCTATAGGGCAGCCAGCAGAGCTCTCTACA	229850
QY	1240	ysGlyThrIleThrArgIlelleGlyGluAspSerProSerArgLeuAspArgGlyArgG	1260
Db	229849	AGGTTACCATCAGCAGGATCGTTGGTGAGGACAGCCAAAGTCGCTTGTATCGGGCAGAG	229790
QY	1260	luAspSerLeuProLysGlyHisValIleTyrCluGlyIleValIleGlyHisValIleuSert	1280
Db	229789	AGGATACCTTGCCCAAGGGCCATGTCTATGAGGGCAAGAAAGCCATGTCTCTATFCT	229730

QY	1280	yrGlu-----	1281
Db	229729	ATGAAGTGGGTACAGAGGACAGAGACCATGAGGCACTGGCAGTCAGGTGGGTGGCTCC	229670
QY	1281	-----	1281
Db	229669	TGCTGTCCCTTTCTTCTTCCATTCTAAGGATGTCTTTTCTCAAGACAGACATAGCAAGACT	229610
QY	1281	-----	1281
Db	229609	CCTGAGAGCAATCCCTTCTGCTAGGCTGGGCTCAGCAGGACATGTCAGACAGTACCTCC	229550
QY	1281	-----	1281
Db	229549	CTCCAGGTCAGGTTCCTTCTGCAAGGAGTGTGTGGAAACCCCTTGTGTAAGGAGTC	229490
QY	1281	-----	1281
Db	229489	AGGAAAACAGACTGTAAACATGTGTCCCAAGACAGCAGTGTGGTGTCTATACACTTGC	229430
QY	1281	-----	1281
Db	229429	TGAGAGGGAAGTCTTTTGGCAGAGCAGAAAGCAGCCAGGGGCTTTGGGTTGGTGTG	229370
QY	1281	-----	1281
Db	229369	TGCTAGCCTCAACAGGTTGAGGACTGGGCTAGGCCCTGCTGCCACTTAGACATAGGG	229310
QY	1281	-----	1281
Db	229309	AAACGGGAGATGGACATAGTCATTCAGTGGCTTCTATGAGGATACCATCTGGGCC	229250
QY	1281	-----	1281
Db	229249	TGCTTAGACAAGCAACCTTGTGTAGGAGAGCCATGGATCTATAAAGTGTCTTT	229190
QY	1281	-----	1281
Db	229189	AGACAAGGCAGAGCTGAAGGACACCAGGAGCAGGAGGAGTGACTTCTCTCTCTGC	229130
QY	1281	-----	1281
Db	229129	AGGCATACAGCCAGAGGGGTGAGGGCTGGCTTGGTGGCTTCCCTACAGAGGTGGG	229070
QY	1281	-----	1281
Db	229069	AAGCTGGGGAATCTTTGAGTGGCTGACAGGAAAGGGCTGTGGTAGAGCCATGACACC	229010
QY	1281	-----	1281
Db	229009	TTCTCTCTGGCTAGAAAAGGAGCCAGGGCTGCCCTAAAAGTTGAGGAACCCATACCTGGGA	228950
QY	1281	-----	1281
Db	228949	TGCGTCTCCAGTACCAAGACCCACATCAACTGCTGCCATGACTACTGAGGCGCTG	228890
QY	1281	-----	1281
Db	228889	GGGGCTCCAGGTTCCCTATTCTTGTAGTCTCAGAGCCCTGGTCTCTAGAGTCAAG	228830
QY	1281	-----	1281
Db	228829	TGGCCCCATGGCTTTGGCACCTGACAATAAGATACAGTCTTAAGGACAGGGCTACAGC	228770
QY	1281	-----	1281
Db	228769	TGACCCCGGTGCTGGTGCATGTCCAGGTGACAGACCCCAAGAGCAGAGAGCGGTGAT	228710
QY	1281	-----	1281
Db	228709	CAGAGTCTTGGCCAGAGCCCTGAGTCACGAGTGGAGGCTCTAGAGTCTCAGGTGTGA	228650
QY	1281	-----	1281

Db	228649	ACTTTACAGCCTATCTTGGTGGAGACTGGGTATCCTGTGTCTCTTTGGGGGCCAGTGG	228590
QY	1281	-----	1281
Db	228589	GTCTGAAGCAGTTTCAGAGTAGCGCCCAACCGTCAGGCCAGCCAGCTGCCAGCCCTTTC	228530
QY	1281	-----	1281
Db	228529	TTGCGCAATCTCCATAGCCAGAGCATCCAAGTCATTGGGGATAGCAATGCCACTCTGTCC	228470
QY	1281	-----	1281
Db	228469	CGACACGCTGAGGCTTTCATGGTCTTGGACCCGGTGTCAACATCCCTTGTGTACA	228410
QY	1281	-----	1281
Db	228409	TGCAGACCTGCAGGCGCTTGTGGGGCTGCTCTGCGGACTCAGACCGTGTGACATCTTC	228350
QY	1281	-----	1281
Db	228349	GTTAGTGTGCACATACACCATGGGCACAGACACATACACCCCTACACAGCAGTGG	228290
QY	1281	-----	1281
Db	228289	CAGGCATGCTGGCACTTTCTGCCAAGACAGTTCCTCAGTGTGAGTGGTGTGAGGTTCC	228230
QY	1281	-----	1281
Db	228229	ATCTGGGGGCGAGGGCCTCTTTTCTGCTGTGCTTTCCTTCACTCACTGCTGTGGT	228170
QY	1282	-----	1285
Db	228169	GTCTGCACCTTGTGTCACCCACACCATGTGCTTTTCTGGCCCTCAGGTGTGTGTGG	228110
QY	1286	ValThrGlnCysSerlysgluaspGlyArgSerSerSerGlyProProHisGluThrAla	1305
Db	228109	GTGTCAAGTGTCTCAAGGAGGATGGAGGAGCAGCTCAGGCCACCCCAAGACTGCC	228050
QY	1306	AlaProLYsArgThrTyrAspMetMetGluGlyValGlyArgAlaIleSerSerAla	1325
Db	228049	GCCCCTAACGCACCTATGACATGATGGAGGCCGTGTGGGAGGACCATTCACTCAGCC	227990
QY	1326	SerIleGluGlyLeu-----	1330
Db	227989	AGCATCGAGGGTAA-GTGTCTCACTGTAGTGTATTACCGGGGTCAAGCACCATCCAGGA	227931
QY	1330	-----	1330
Db	227930	GACCCAGGTACCCAGGTAAAGATGGAAGGGAGGGTGGCCCTTGACATGCTGGAGGG	227871
QY	1330	-----	1330
Db	227870	AGCCAGTCAGTAGTGGCTCAATATTAAACAGTAGAGAACTAGGACACAACCCGGTGCA	227811
QY	1330	-----	1330
Db	227810	AAATGAGCCCGAGGAAACGGGSCACTTGTCTTGTGGTGTAAAGTCACAACAGAGACTG	227751
QY	1330	-----	1330
Db	227750	TTGGTGTAGAGCCAGCCCTCTTGACACCATACGGTGCACACTATCCCTGTCTCCCCA	227691
QY	1331	-----	1348
Db	227690	GGACTCATGGCGGTGCCATC---CCTGACGACACAGCCCC---CACCTCAAGGAGCAA	227637
QY	1349	HisHisIleArgGlySerIleThrGln-----	1357
Db	227636	CATCATCCGAGGCTCCATCAGCAAGGTATATCTCTTACGGTGTGTAAATCCACCT	227577
QY	1357	-----	1357

Db	227576	CCAGCCAAAACCAAGCTTAAGGGGTGGAAACCGAGGGTGAACCTCCGTCCTGTCTTAT	227517
QY	1357	-----	1357
Db	227516	AGGTGGAGTCACTACTCTGTACTTGGGGCTCATTCMAAGCCCATGCTGCATGCTTGCT	227457
QY	1357	-----	1357
Db	227456	AGTGATTGAGAGGGGTCCAGATTGAATGTCTTTAGTCTCATGCTTTGTGTAAACTG	227397
QY	1357	-----	1357
Db	227396	AGCAATGTGTTTCACTTAGCTCTCACTCATATGAAGAAACAGGCTTAGAAAAAGTCA	227337
QY	1357	-----	1357
Db	227336	GTTTGCATACCCAAAGTCCACAGGAAGATTGTGTGACCCAGGCTGAGCCAGGCAAG	227277
QY	1358	-----	1367
Db	227276	GCAGTGGAGAGCCAATGATGTTTCCCGGCGAGGCATCCCGAGGTCTATATGTGGAGGCCA	227217
QY	1367	nGluAspTyrLeuArgGluAlaLysLeuLeuLysArgGluGlyThrProProPro	1387
Db	227216	GGAGGACTACTTACGTAGGAGGCCAAGCTTTGAAGCGGAGAGGCAAGCCACCGCCCC	227157
QY	1387	oProProSerArgAspLeuThrGluAlaTyrLysThrGln-----AlaLeuGlyPr	1404
Db	227156	ACCGCACCTCGGAGCTGACTGAGACCTCAAGACCCGCGCCCTCGACCCCTTGGGTCC	227097
QY	1404	oLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgse	1424
Db	227096	CCTGAAGCTGAAGCCAACTCATGAGGGTGTGTAGGACTGTGAGAGGCGGGCGCTC	227037
QY	1424	rIleHisGluIleProArgGluLeuArgHisThrProGluLeuProLeuAlaProAr	1444
Db	227036	TATCCATGAGATCCACGAGAGAGTGTGCGCGACGCGCGAGCTGCCCCCTGGCACACG	226977
QY	1444	gProLeuLysGluGlySerIleThr-----	1452
Db	226976	GCCTCTGAGGAGGGTTCCATCAC-CCAGTACGAGGAGCACAAAGAGGGCGCGGGCT	226918
QY	1452	-----	1452
Db	226917	TGGGCAAGCCTGAGGGAGTTCTAATGCTGCGCTCATTTGCATTGGTTATATAAATAAGGCC	226858
QY	1452	-----	1452
Db	226857	CAGAGAGGTCAGCGGCTTGCTCTGCTCACTGACAGGGCTTTGTGAACCAAAAGTCTTTTGG	226798
QY	1452	-----	1452
Db	226797	TCATTGGCTGCTCCCAACATGCTACCTGCTGTTGTGTGTGTCAAGACTAGAGCCACTGT	226738
QY	1452	-----	1452
Db	226737	CCCTACCTTGGTCATGGCTTGATGGAAAGAAAGACAAACAGCCAGGACACCTA	226678
QY	1452	-----	1452
Db	226677	GGGAAGGCTGTCTGTACCTGAGCCAGGGACACACAGTGTATGGACCAATGATTTCTAGG	226618
QY	1452	-----	1452
Db	226617	GTATCTTTTGAACCTTAGGTGCTGCTCCTCATTTGGGCCAGCACTTGGGAGGGGAG	226558
QY	1452	-----	1452
Db	226557	TCAGGAAGATTGTAATTTGAGGCTAGCTAGGCTACAAAGCTGTACTCTGGCTTACTGC	226498
QY	1452	-----	1452
Db	226497	CGATCACCMAAGTTTTACAAATTAATGACAAACAAACATTGAGTGAAATTGAGATCTGG	226438
QY	1452	-----	1452
QY	1452	-----	1452
Db	226437	GACCTAGCTGTGTCACTGTCCCGAGCCTGGATCCCTGTGGGATCCTCTCCAAGATGTCC	226378
QY	1452	-----	1452
Db	226377	AATAATTCCCGTCTCCGTGGTGTCTCTGAGAGCAACAGGAAATGTAACTCCAGCCT	226318
QY	1452	-----	1452
Db	226317	GGACCCGAGCAGAGCTGGGGTGAACAGATGCAATCTGATATGGAGTCCATAGTGCT	226258
QY	1452	-----	1452
Db	226257	GCAGGATGTCGGGGGCCACACTGGGTGAGGAAGTCAATTGACCTGTAGTCCACGC	226198
QY	1452	-----	1452
Db	226197	TCTCGTACCCCTCCACAGTGGGGTCTCTGGACGTATAGCCAGGCTGAAGCGAGCAGG	226138
QY	1452	-----	1452
Db	226137	TATCTAGCCTCTGACGTCTGTCTTGTCAAGGGTTAGAGATGCCCATCTCTGTGCAGAC	226078
QY	1453	-----	1455
Db	226077	ATACAGGAGAAACAGCATCTGGCCAGCCACATACAGACCTGGACCTCTCCCGAGGGAC	226018
QY	1455	rProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgse	1475
Db	226017	CCCACTCAAGTAGCACTCCGGGTGCCCTCCAGTGGCTCAAGAAACACGACGTACGCTC	225958
QY	1475	rLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaas	1495
Db	225957	CATCATCGCAGCCCGCGCGGCTTTCTCTGCCCTGCACCCCACTGGACATAATGGCTGA	225898
QY	1495	pAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAl	1515
Db	225897	TGCCCCGGGACCTAGAGCGGCTGTCTATGAAGAGCTCTGAAGAGCCGTCAGGACCCAG	225838
QY	1515	aSerSerSerGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLy	1535
Db	225837	CAGTGTGTGAGGGGGTCCATCACACGCGGGGCCAGTTGTCTGTCTGTCTGTCTG	225778
QY	1535	sProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuPr	1555
Db	225777	GCCCCGGCAAGGCCCACTGGCTTACGAGGACCAACGCGGGCACCCCTTCCAGCCACCTGCC	225718
QY	1555	oArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSe	1575
Db	225717	CCGAGGCTCACCTGTGACTACCAGGAGGCCACGCCACCGCTTCAGGAAGGTGAGGC-TC	225659
QY	1575	r-----	1575
Db	225658	CCCTGGGCTGCCAGCTGTGGGGAGGGGACAGCTCATTTTCCACCGCCAGTCTGTCTGTG	225599
QY	1575	-----	1575
Db	225598	CCCTGCCCTCCACACTGCATACAGAGCCCCCGTGTCTCCCTTACCAATATATTTTAA	225539
QY	1575	-----	1575
Db	225538	TAATTCMAATCAGGGCTGAGGAGAGGGCTCAGTAGTAAAGTCTTCTGTACCAGCATGA	225479
QY	1575	-----	1575
Db	225478	GACTCAGAGTTCAAATCCCGGTACCCACATTAAGCCAGGACACACTCATACACATCTGTAA	225419
QY	1575	-----	1575
Db	225418	TTCTCATGCTTCAAGTGAAGCAAAAGGCTAAGACAGAAATCTCTTAAAGTTTAGGCCA	225359

Qy	1575	-----	1575
Db	225358	TGAGTCGGGGTCTCAAAACATGGCGAAAGTGGATGACACCTCTCCCCCCCCCTCTATT	225299
Qy	1575	-----	1575
Db	225298	TTATACACGCATACGACACACTCATATCAGCACACTCGCGCCTACATTCCTGTGCAC	225239
Qy	1575	-----	1575
Db	225238	ACATGTACACAAAAACAATAGTTTAAATGAAGCAGTTCACACTCTGTAGGCTTCACCT	225179
Qy	1575	-----	1575
Db	225178	CTCCAGTAGGGCTGCTGGTCTGAAGCACCTGCCTGTGACGTGACCGCGCATCCTACTCT	225119
Qy	1575	-----	1575
Db	225118	GAATCCGCTTCTGATTCCTCGCATTCCTGACCTCCGCTCTTTCGATCCTCCACAT	225059
Qy	1575	-----	1575
Db	225058	GTGCAGCCTTTCGCTGTGGCCCTCTATGCCGAAGGCTCCACCTTTGGTTTGTCTCT	224999
Qy	1575	-----	1575
Db	224998	AGCCAACCACTGTATACCCCAACGTCACAGATGTCAGGGCTGAGAGGCTATAGGCACC	224939
Qy	1575	-----	1575
Db	224938	TCCAACATGTTCCAGCCACATPCGGGGCGGGATGTTGTTAGGGTGTGACACCATGTTTT	224879
Qy	1575	-----	1575
Db	224878	AATTCATGTTATGAGGAGCCCTCCAGAGCACTGAGATCCTCAACATCTCTCTCTCT	224819
Qy	1575	-----	1575
Db	224818	CAGGCCGCTGCACCTTCTACCTTGGGTTTCATCTCTATCTGGAACAATGGCTCCTGGAG	224759
Qy	1575	-----	1575
Db	224758	CTCAGGGTCTTGCTGCTCTTCTCTCAGGGTGTCTCCTCTCTCATGGGAATGCC	224699
Qy	1575	-----	1575
Db	224698	AGCCCATGATTTAAGTCAGACTGGAATGAATAGGGCAATGGTGNCTGGGAGTAAATGAA	224639
Qy	1575	-----	1575
Db	224638	TGCTTTTCAAGAGGACGAGGACGAGAGACAAGAGGCTCTAGGGCTTGTGACAGCTCAG	224579
Qy	1576	-----Se	1576
Db	224578	CCTGGTGGCCGCTGAGGCTGAGGCCACTTTGACCTGTCCCCACAGGCACTCTCTCTC	224519
Qy	1576	rSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerPr	1596
Db	224518	CAGCAAGGCGTCCAGAGACCGGAAGCTGACCTGCACACCCCGGGAGATTGCCAAGTCCCC	224459
Qy	1596	oHisSerThrValProGluHisProHisProHisProHisProHisProHisLeuLeuAr	1616
Db	224458	ACATAGCACTGTGCCGAGCACCAACCCACCCCATCTCCCCCTACGAGCACCTGCTCCG	224399
Qy	1616	gGlyValSerGlyValAspLeuThrArgSerHisIleProLeuAlaPheAspProThrSe	1636
Db	224398	GGGTGTGACTGGTGTGACCTGTACCGTGGCCACATCCCATTTGACCTTTGACCCCACTC	224339
Qy	1636	rIleProArgGlyIleProLeuAsp-----	1644
Db	224338	CATACCCCGAGGATCCCTCTGGAAGCAGGTGGGTGCTGGGCTGTACATGTGGGGCTT	224279
Qy	1644	-----	1644
Db	224278	ATTGATTGTCTCTCTGGGTGGACACTTGGGGTATGGCTTAGTGTGCTGCCATCACCAGCT	224219
Qy	1644	-----	1644
Db	224218	GCTTACCCAGGGCAGGTTCTAGGATTTGGGCTTTTTCGCGTGTCTCACCCTAGGACT	224159
Qy	1644	-----	1644
Db	224158	CATCTGTCCCTTGCAGTGCAGTAAATGAGGGGCTGCAGTGAGTGCCATGATGTGGAACC	224099
Qy	1644	-----	1644
Db	224098	TTTTTAAAGGCTCTCTTTTGAAGGTGTAGCACCTGTATTTTCGATGCATTATATTGGGG	224039
Qy	1644	-----	1644
Db	224038	GCATCTCCGAAACATTAAAGCAGGCCCCCAACATGATGAGGGAGCGAGTCTTTAAGCTC	223979
Qy	1644	-----	1644
Db	223978	TGTGTGTGCATTGGGGAGGCATGTGGGTGTACAAGGAGAGGAAGCTGTTGTGGCAGAAGT	223919
Qy	1645	-----Al	1645
Db	223918	GAGGAACGTGAAGATAAGTCATGGGCCACCTGACCTCCATCCCTCCCTCCCTCCCGAC	223859
Qy	1645	aAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTy	1665
Db	223858	AGCCGAGCCTACTTACCTGCCCGGCACCTTGGGCCCCCAGCCACCTTACCACACCTGTA	223799
Qy	1665	rProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIl	1685
Db	223798	CCGCGCTTACCTCATCCGCGGCTACCTTGACACCGCGGCCCTGGAGAACCGCCAGACAT	223739
Qy	1685	eIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGl	1705
Db	223738	CATCAATGACTACATCACTCGCAGCAGATGCACCAACGAGCCTCTGCCATGGGCCCA	223679
Qy	1705	nArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAl	1725
Db	223678	GCSCGCGATATCTGAGGGTCTGTACCCGAGAGATCTCTCGTGGCCCTCAATTATGC	223619
Qy	1725	aAlaGlyProArg-----	1729
Db	223618	CGTGGCCCGCTGGTGGTGGCTGGGCTGTGCCCATCAGCAGAGACTTGTCTCTGGC	223559
Qy	1729	-----	1729
Db	223558	TGTGGTGTGCAAAAGAAGTGCACACAGTGCACCTGCAGCTGGGCGGAGGGGCACAGTT	223499
Qy	1729	-----	1729
Db	223498	GGCTGGAGGACTAGTGGTGTGCATGCTGGCTCTTGGGGAAAGGGGACAGCGGGCCACTT	223439
Qy	1729	-----	1729
Db	223438	GTATCTGTTACATCTAGAACTGTGCTTAGCAGACACTCTTGGCCACCATGCACCGTGGT	223379
Qy	1729	-----	1729
Db	223378	CTTTGGCTGTTTCTGAAACCTTGGTATGGGCACTCTTCTGTATCCAAAGAGATCAAAA	223319
Qy	1729	-----	1729
Db	223318	TACAGCTGGATGTGCATTTGACGCCAGCGCTATATCTTAAGTACGAGAGAGGTGCAAGGCCAG	223259
Qy	1729	-----	1729
Db	223258	CCTCAGCTACATAAAATTTTGGAGGCCAGCCTGGGCTATATAGTCTGCCTCTATACAAATG	223199
Qy	1729	-----	1729

Qy	1898	-----	1898
Db	221009	CTTCTCGCTTCCACTCAGCATGCTTATATCTACATGTCATGTAGCAGTGGATGCCCTC	220950
Qy	1898	-----	1898
Db	220949	AGACGAGATGGCTGCTCTGTCTGTCTATCTCAGAAGCTCCGGACCCCTTGCCCTTAAT	220890
Qy	1898	-----	1898
Db	220889	TACCAAAATTCGATCATCTGGAGGGTCTTGGCACCTGGTGGAGCTCAGTAAAGACT	220830
Qy	1898	-----	1898
Db	220829	TTTGGGCCAAGCACAGCTTAGGTAGTCAGTCAGGAGACACTTATTCTGGGAGCCTCAGGGAG	220770
Qy	1898	-----	1898
Db	220769	TGAAAGTCTAGTAGGAACATGGAGGTGTGAGTAGGAATTCAATTGGCCAGCACTGCCT	220710
Qy	1899	-----	1912
Db	220709	GAACCTTCTCTTCTTTTTCAGTCCACCTCCACTCTTCCACCTGTCCGCCAGCTGCCA	220650
Qy	1912	hrPheProAlaIaThrHisCysProLeuGlyThrLeuAspGlyValTyProThrL	1932
Db	220649	CAITCCACCTGCCACCACTGCCCCACTTGGTGCCACCTCGAAGGGGTCTACCCATACC	220590
Qy	1932	euMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProAla	1952
Db	220589	TCATGGAGCAGTCTCTGTACCCAAAGGAGACCTCTCGGGTTACCCGGCCTGAGCGGCC	220530
Qy	1952	rgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProA	1972
Db	220529	GTGTGGAGCTGGCCATGCTTCTTCCACCAAGCCCCAGCCCG-----GAGCTG	220479
Qy	1972	laSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisA	1992
Db	220478	CCTCTCACCAGCAAGAGCTCCGAACCCGATCCCTTAGCACCCCCAGCTCCAGCCACA	220419
Qy	1992	laThrIleAlaArgThrProAlaLysLeuAlaProHisHisAlaSerProAspProp	2012
Db	220418	CAGCATCGCCGACCCAGCAAGAGCTTGCACCCACCATGCCAGTCCGAGCCTGC	220359
Qy	2012	roAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheS	2032
Db	220358	CGCG-----CCCACTCGGCTCAGATCTGCACCGAGAAAGACTCAAAGTAAACCCCTTT	220302
Qy	2032	erIleGlnGluLeuGluLeuArgSerLeu-----	2041
Db	220301	CCATCCAGGAATTGGAATCCGTCTCTGGTAAAGACCACCCTGACAGCGGCCACGTTCA	220242
Qy	2041	-----	2041
Db	220241	TAGACGCAATATACCGCTCAAAATTCTCAGAAAGGGGGCCGCGAGAGGAGTTGCG	220182
Qy	2041	-----	2041
Db	220181	TGGCCATCGACTCCCTGGCGCAGGTAAAGACACCCGCGCCGCTGCCACCCCATCT	220122
Qy	2041	-----	2041
Db	220121	GTGSCCTAAAGATATTTTCAGATCTTGTCTTTTAAATTTTTCGCCCTTTTGTGGTT	220062
Qy	2041	-----	2041
Db	220061	TTGGTATTCGTTTGTAGCTCGCCATCTCTCAGTCGGCAGCGTGTGTGACTCCGCCC	220002
Qy	2041	-----	2041
Db	220001	ACAGGGCCCTCATTCGCCCATCTTCATCATGAGTTTTTTTTTTCCTTTTCTT	219942
Qy	2041	-----	2041

Db	219941	TTTCTTTTTTTTTTTTCTCTCTTTTCTTTTGGATTTTTCGCTTTTTCCTTTAATG	219882
Qy	2041	-----	2041
Db	219881	AATGGATCTGTGTTTGGACTCCGACCGCGCTCCCATCTCCCGCCACACCACTGCAC	219822
Qy	2041	-----	2041
Db	219821	CTTCGTTTGGGAGTGGGATGCAGGGTCTACAGTTGACTGTGCCACCACTGTACATG	219762
Qy	2041	-----	2041
Db	219761	GAGAAAGTTAAGGCAGGAGTGTGGTGGGGGAGGCTGGGTACCCACTCTCTTCCCA	219702
Qy	2041	-----	2041
Db	219701	CACAGCTACAGCCAGCCCTAAGGGAATGAAGCCTCCCGCTGGATCAGAGACAGTGGGA	219642
Qy	2041	-----	2041
Db	219641	CTGGGGCAAGGCTGGACCTTAGGAGCCCATTTGTCACTCAGCTTGAGCCTCACACCAACCAC	219582
Qy	2041	-----	2041
Db	219581	ACCTGACCCCATGCACATTCACACCCCTCCCTCCAGTTTCTCCACCAACCACCTCCAGTC	219522
Qy	2041	-----	2041
Db	219521	TCCCTATCTGAGCCCCACACCGCCCAAGTCAGACTCCCATTAATCTTTGGGTGTCTG	219462
Qy	2041	-----	2041
Db	219461	AGTCTCATATAATGGAGCTCCCTTCACTTGACAGAGTTCCTCTCTGSCCACCCTAGGTC	219402
Qy	2041	-----	2041
Db	219401	GACCTGATTTTCTTCATCTTGAGGCCAGACTGTGGAGGAATGGCTCTCTCCCAAGCCCA	219342
Qy	2042	-----	2050
Db	219341	CCCCACATTTGACCTTAGCTCACCTGTCTCAGGTTTACCACAGTGGAGCTGGCTACAGCC	219282
Qy	2050	roGluGlyValGluProValSerProValSerProSerLeuThrHisAspLysGlyL	2070
Db	219281	CTGATGGGTAGAGCCCATCAGCCCGTGAAGTCCGCCACCTGACCCACGACCAAGGGG	219222
Qy	2070	euProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysG	2090
Db	219221	TCTCCAAGCCTCTGGAAGAGCTAGAGAAGACCACTTGAGGGGGAGCTGGGCACAAGC	219162
Qy	2090	lnProGly-----	2092
Db	219161	AGCCAGGTGACACCCGGGAGAGTAGCAGTGGGGAAACTGAGGACGGCTCCATGTTGTC	219102
Qy	2093	-----	2100
Db	219101	CTCAGGCCCTGACAAAGCTCTGCTTGTGTCTCAGGCCCCATG-AGCTCAGTGGGGAGG	219043
Qy	2100	laAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerProLeuL	2120
Db	219042	CTGCCACCTCCACATCTCGCGCCACTGCGGAGAGCCAGCCCTCATCCAGGCCACTCC	218983
Qy	2120	euGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleS	2140
Db	218982	TCAGACTGCCCGCAGGCATCAAAAGGTCAACAGAGGGTGGTCACTTGGCTCAGCACATCA	218923
Qy	2140	ex-----	2140
Db	218922	GTGTAAAGTCCCATCTCTGCCCCCTGCCATTGGGAAGGAGGTGGCTGTATGACAGCCATAT	218863
Qy	2140	-----	2140

Db	218862	CCACAGCCAGACAGATGCTTAACGTAACTGGCCACCAGCTTAGTCTTTGGGGT	218803
QY	2140	-----	2140
Db	218802	CATCATAGGTCTTAACAGGGACACTTAACAGGAGACCAAGAGCCAGATCAGGGTTCTG	218743
QY	2141	-----	2146
Db	218742	CCCTACTACTCTGAAGTCTCTGAACTGCTCCCTCTCCCTGAGGAGGTCAATACAGAGAC	218693
QY	2147	TyrThrArgHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPhe	2166
Db	218682	TACACAGCCACCAACACAGCAGCTCAGTGGCCCCCTTCCGCCCTCTCTACTCTCTC	218623
QY	2167	ProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuPro	2186
Db	218622	CCTGGAGCCAGCTGCCCTGCTCTGGAATCTCCGCCCTCCACCCAGTGACCTCTACCTCCCA	218563
QY	2187	ProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyLys	2204
Db	218562	CCCCCGACCATGCGACCCCGCCGAGCCCGGGATCCCCCACAGTGAAGGGGGCAAAAGGTGA	218503
QY	2204	-----	2204
Db	218502	GGAAGTACCCATGCTGTCATCAGGTGCTGGGAACAGTGGGACAGGGGTGAGCTTTGCAGT	218443
QY	2204	-----	2204
Db	218442	GTGGGCTGATGGAAACTGAGGTGAGGGATACAAAGCCTCTACGCTCCAGAACAG	218383
QY	2204	-----	2204
Db	218382	ACAAAGCATCTAATCAAGAAGTGGCTCACCAAGGGTCCACCGTGCCAAATTGAATGT	218323
QY	2204	-----	2204
Db	218322	CCTATATACCAACTCAGCTAGGGGTGAGATGTCCCACTTCATTTAGCAGATAGGAA	218263
QY	2204	-----	2204
Db	218262	ACTGAGCATPAGAGGTGAGTTACTATCCCGTCCCATGCACAAACACCTGTCACAGA	218203
QY	2204	-----	2204
Db	218202	TGTAGACAGGTATCTCTGGGTAAACAGCGAGGTCTTAGCATGGGGCTCATCAGGGCTTA	218143
QY	2204	-----	2204
Db	218142	CAACATGGCTCCAAGCACCCAGCAGCACTGTGTGTGGTGGCATACCTCTGGACAGGCTGA	218083
QY	2204	-----	2204
Db	218082	GGCAAGACGACTGTGAACCTGAGGGCAGCGCTGGGTGTACCAAGACCCCTGACTCAAAA	218023
QY	2204	-----	2204
Db	218022	ATAAACCAAGCCATGAGTTGGGTGAGTTGGAAGGTCTCAGAAAGCTCTCAGAAAGCTCTGGGGTC	217963
QY	2204	-----	2204
Db	217962	TCCATACCTCAGATGCAGAGTCTGGCCATTTACAGGGAAACAGGTGAGCCCCCTGTGGG	217903
QY	2204	-----	2204
Db	217902	CAGTAGGGGTGATCTTCAATGGAGTAGAACCCCACTCTCAGTGGGATGCTGTCTCCATT	217843
QY	2205	ArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGlyIleGluPro	2224
Db	217842	AGGTCTCCAGAACCCACCAACA---GTCTGAGTAGCAGCGAGGATGCTATTGAACCT	217786
QY	2225	ValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeu	2244
Db	217785	GTGTCCCCACGGAGGGCATGACTGAGCCAGGACATGCACGGAGCACTGGGTATCCCACTG	217726
QY	2245	LeuTyrArgAspGlyGluGlnThrGlu-ProSer	2255
Db	217725	CTGTACCGGACGGGAGCAGCGGGAGCCAGGTAGTATGSCACAAATCCCGCTTAGA	217666
QY	2255	-----	2255
Db	217665	GGTGGTGGTGTGGTTCCCATACACCCGCTGCGCTCAGCGCTTAGCGTGGAGTGTCTCA	217606
QY	2255	-----	2255
Db	217605	ATCTCTTACAGCTCAGCGGGGTGATGACTTTTCAGGGGGCTTTCAGTGGGACCCCTGTG	217546
QY	2255	-----	2255
Db	217545	CCTGGCATCTCAGCCGAGTCATGTCCTATTTATCATTGTGCTGATTGAGAATGAACCT	217486
QY	2255	-----	2255
Db	217485	AAACGGACACAGAGTGGGTAGGGTGGGCATGGAACCTCAGGAGGGGTATGACCTATGGCG	217426
QY	2255	-----	2255
Db	217425	GCCAGTCCCTGGGAATTTCTGTGTCAGTATAGCTCTTAAGAAGTTCTCAGGGGGAATTC	217366
QY	2255	-----	2255
Db	217365	ACAGTGTGCTGGAGGCTCACAAACCGCCCTAAGCCAGGGCTAAGTACAGATCCACGCA	217306
QY	2255	-----	2255
Db	217305	GGCTGGGATGGAAATGTTGTTCTTAGCACCTCACTGTGGCTCAAGTGGTGGAGCTAG	217246
QY	2256	-----	2256
Db	217245	CCCCCTGTTGCTAAACACAGTCCCGAGCCTGTGACCTCGGCTCTCATGCTCCCTGCGAG	217186
QY	2256	gMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuTh	2276
Db	217185	GATGGTCTTAAGTCTCCAGGCAACACACGAGCCAGCCGCGCTCTTCTCAGTAAAGTGA	217126
QY	2276	rGluSerAsnSerAlaMetValLysSerLysGlnGluIleAsnLysLysLeuAsnTh	2296
Db	217125	CGAGACAACTCCGCCATGTTGAAGTCGAAGAGCAGGAGATCAACAGAAACTCAACAC	217066
QY	2296	rHisAsnArgAsnGluProGluTyr	2304
Db	217065	CCATAACCCGAATGAGCCTGAATACAGTAAGAGAGAACCAAGCCAGACGAGCACTGGG	217006
QY	2304	-----	2304
Db	217005	AGACAGAGTGTCTTTAAGCTTCTGTCGAAGGGCTGATGCTGTCAGTGGATACCGA	216946
QY	2304	-----	2304
Db	216945	CGGCATATAGCACCTCAGTAATGCCCTACGCACAGTAGGTGCTCAGCGAAACCTATAGAG	216886
QY	2304	-----	2304
Db	216885	ACAGCCCTGTGTGGGCGGAGGACAGTGCATCTACTCCAGGTTTCATGCTGCAGGCAT	216826
QY	2304	-----	2304
Db	216825	GGCGCCTTTTCCACAGCTGATGTGGTGGACAAAGGACCTTGTCTTAGCTCAGGGTCG	216766
QY	2304	-----	2304
Db	216765	CAGGACCATATGCTTTATGAGAATGCACTGGGAAACAAACATTTGAGGGCTTCGCACTGT	216706
QY	2304	-----	2304
Db	216705	CCCTCCACAGCAGGCCCTCCCTAGGTGGCTTGGACAAATCCTTTCTGGAGTTAGACT	216646

Db	214486	ACCCTGGCATTTCGAGAAAGTCCACAGACACAAATTCATGAGAGGAGAGAAAGAAACCG	214427
QY	2322	-----	2322
Db	214426	GAGAGCCTTTGAGAGGCTCAGTGGAGGTGCAACATCTTTGAGGGCAGTGTGTGAT	214367
QY	2322	-----	2322
Db	214366	TTGACTACTCTCTCTGTGTCATGGGGTTGGGCCAGCCAGGGCTGGAGTGGCCCTTCACCT	214307
QY	2322	-----	2322
Db	214306	GTTAAGCGCTACCTGTCTTGACGCACATAACTGCATGGTGTGACCCCCCATCCCT	214247
QY	2322	-----	2322
Db	214246	TGTACCCAGCTCTACCTTTCTCTGGCTCTGTGGGAGAAATGGGGTGGAGGGGGCGAG	214187
QY	2322	-----	2322
Db	214186	CCTAGCTAAGGGGGAATCAGGAGGCTGAGTCCCTAACCCCTGAGAAAGGGGAAACNA	214127
QY	2322	-----	2322
Db	214126	CAGGGGCTCTCTCTGGTCACTTTCTGTCCTCCACATCCCTTCTAGAACCTACTCACAG	214067
QY	2322	-----	2322
Db	214066	AAGACCCGGGTGGAATCTATATCCCGTCCCTAGCACCTTAGCTGTGGTTAACTCCATAG	214007
QY	2322	-----	2322
Db	214006	CACAGCTGCCCTGCCTCTTCTTAGACATCACCTTATAGTCAAAATGGAATAATAGAC	213947
QY	2322	-----	2322
Db	213946	CCAGCTGCGATTGGGGTGTAGAGTAGGAACAGGGCTGGGACCTCCGGCTGCCACAGAC	213887
QY	2322	-----	2322
Db	213886	CCCTCCCTGCAACAGCCATGAGCTGTGGGGCAGTTAGGACCCCATCCGACCTTACAC	213827
QY	2322	-----	2322
Db	213826	ACAGAGCGGGCTGGAGGACGACATGCTCGCAGATGTGTCTCCTCTTTC	213767
QY	2322	-----	2322
Db	213766	TCTCTCTCCCTCCTCATTTCTCTCGCAACCTCTGGCTCCCGGCCCGGCCCC	213707
QY	2323	-----	2334
Db	213706	TGTGTGTGTCTGTCTCTCTCCAGGCTTATGACCTTAGAAGCCAGCGGTGCAAGA	213647
QY	2334	uHisAlaSerThrAsnMetGlyLeuGluAlaIleArgLysAlaLeuMetGlyLysTy	2354
Db	213646	ACAGCCACGACCAACATGGGCTAGAGGCCATTATTAGAGGCACTCATGGTAATA	213587
QY	2354	rAspGlnTrpGluSerProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSe	2374
Db	213586	TGATCAGTGGGAAGAGCCCGCGCTCGCGCCAAATGCTTTTAAACCTCTGAATGCCAG	213527
QY	2374	rAlaSerLeuPro--AlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisTh	2393
Db	213526	TGCCAGTGTGCCGCTGTGCTATGCCATAACCACTGTCTGAGGACGAGTGACACCC	213467
QY	2393	rLeuThrSerPro-----	2397
Db	213466	ACTCACCTCGCCAGGCTCGAGGCTGCCCTGCCACCGCCCGCCGCTGTGTCCCT	213407
QY	2397	-----	2397
Db	213406	CCTTACCCTGGTGTGTGTTAATCCTTTTCACTCCCGCCCTCAAGCTTTGGAGAGACTGTGACC	213347
QY	2397	-----	2397
Db	213346	TTATTTTGTGGTGTTTGACATCTAGTTTGTCTAATCAAAGCTGGTCTGACACCCCATG	213287
QY	2397	-----	2397
Db	213286	TTGCAAGGGGGTCCCTCCCTCGCCCCCAGCACTGTCTTTTCCCGCTGTCTCACACCATCC	213227
QY	2397	-----	2397
Db	213226	CCACATCCACCCCTTCCTCCCTGCTGTGTGGACAGGATCAAGCTGACAGGCCACCC	213167
QY	2397	-----	2397
Db	213166	ATGTGCGACCGGGCTACTGTGTGGGAAGAGTCTAGTGTGGGATGACCCCTTGAAGGCCA	213107
QY	2397	-----	2397
Db	213106	GSCAGCACGAGGCCAACTCCCGTTCTTCCAGTGCTGTGTCAGCACTCTCTGTTCCAAGC	213047
QY	2397	-----	2397
Db	213046	AAACAAGCAAGAGGTGCTCTGTCTCCCTTCTAGGGTCAGCCAGCCTCAAGGTCT	212987
QY	2397	-----	2397
Db	212986	GGCCCTATCAGCAGAGGCTTTCAGTCTAGCACAAATTGAATGCTGTCTTGGGCAATGTCTG	212927
QY	2397	-----	2397
Db	212926	TCCTCCAGCAAGATCTGACACTCTCCTCAGTGAGGGTACAGCTGTGGGGCAGCTAAG	212867
QY	2397	-----	2397
Db	212866	GGAAATAGTCAACAACCTCTCTCATAGCAGATGTTTTTCTTTTTCACGCTAGGGCT	212807
QY	2397	-----	2397
Db	212806	GGCTCTGGCCATCTGGTAGACCTCTTAGCCTCCTGGCTCAGGGTTTCTCACACTAGTTT	212747
QY	2397	-----	2397
Db	212746	CTGTTTGGGGCCTTAGGAGCAGGTGGCTCCCGCAGGCAATCGGCCCCACTGCCCTC	212687
QY	2397	-----	2397
Db	212686	CTCCGTGTGTGTGGGTGACCCACAGTCAACACCATTAACCTGACCCCTTTCCCACTTT	212627
QY	2398	-----	2414
Db	212626	GTCCCTCTGCAGTGTGTGGTGGAAAGCCAAAGTTTCTGGCAGACCCAGCAGCGGAAAGC	212567
QY	2414	alySerProAlaProGlyLeuAlaSerGlyAspArgProSerValSerValHi	2434
Db	212566	CAAGTCGCGCGCCAGCGCTGGCACCCGAGACGACCACTTCTGTCTCTCTCAGTACA	212507
QY	2434	sSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTtpGluAspArgPr	2454
Db	212506	CTCGAGGGCGACTGCAATCGCCGAACCACTCCCAACCGGTGTGTGGAGGACCGGCC	212447
QY	2454	oSerSerAla-----	2457
Db	212446	CTCATCTGCAGTGGGTATCAGCTAGGGCAGAGATGTAGACATGGGGTTCCCTGGGCTA	212387
QY	2457	-----	2457
Db	212386	CCAGGGAGGGGCTCCCTGGTCCCACTAGAAAGAGGGGTAGGAAGCCACACATT	212327
QY	2457	-----	2457
Db	212326	GAGAGGCTCTGTAGATCTCTCAGGCCCTTCAGTCAAGGCTGGTTCACACAGGGGACAG	212267

```
Qy 2457 ----- 2457
Db 212266 TGCCTGTCACGACGTAGGATTTTACAGAGAAAGCAGAGGACGTAGAGGAAAGTTAC 212207
Qy 2457 ----- 2457
Db 212206 CACAGCCGAGTCAAGGTGGTGGTGTCTGTAGCAGAGGCAAGGAGGAGCCGCTT 212147
Qy 2457 ----- 2457
Db 212146 TTGAGAACTCAATGGTGGCTTAGAGGCCAGGCTCGAGCTTCTGTGTGTAGA 212087
Qy 2458 -----G 2458
Db 212086 ATGGGTCTGAGGTGGCTTACTTTCAGAGCCCTCTCCACCACCTCTCCCATCTTTCACAG 212027
Qy 2458 lySerThrProPheProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaT 2478
Db 212026 GGTCCAGCCCATTCCTTACACCTTTGATTATGAGGCTACAGCGAGTGTCTATGGCCT 211967
Qy 2478 erProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaT 2498
Db 211966 CCCACCCACCTGGCTTCCCGCAGCAGCGGGCCCTAGCTGTGTCACCATGCTT 211907
Qy 2498 rpAspGluGluProLysProLeuLeuCySerGlnTyrGluThrLeuSerAspSerGlu 2517
Db 211906 GGGATGAGGAGCCCAAGCCACTGTGTGTTTACAGTACGAGACACTCTCGCAGCGAG 211848

RESULT 16
AC136560/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-128N8, WORKING DRAFT SEQUENCE.
ACCESSION AC136560
VERSION AC136560.2 GI:25188528
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 218129)
Muzny,D.,Maré, Metzker,M.,Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Ayogi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Blawalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,B.,
Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Devila,M.L., Davis,C., Davy-Cartoll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gabregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulsegheh,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,D.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
```

FEATURES

Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajda,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,G., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,J., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

Unpublished
Direct Submission
2 (bases 1 to 218129)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (06-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 218129)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (23-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 23, 2002 this sequence version replaced gi:24635574.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPVF
Center clone name: CH230-128N8
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 208938 bases at least Q40
Consensus quality: 210339 bases at least Q30
Consensus quality: 211081 bases at least Q20
Estimated insert size: 211417; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 218129: contig of 218129 bp in length.

Db	79385	TGTTGCCAGGGTGGTCTGGCTGGACTCAGCCCTGCTGTGTGGGCTACGTGTGCACA	79326
Qy	1126	-----	1126
Db	79325	GGCGTGTGTTCAGTGCCTTAGGCAGGTGACTCACTGTCTGTAGGCTTCAGTTTCATCCT	79266
Qy	1126	-----	1126
Db	79265	CTGTTCACTGGAGGGTGGGAGCTGCAGCTGCTGCAGGCTGTAGAGTGTCACTGTCTGTC	79206
Qy	1126	-----	1126
Db	79205	GTAAGCCCTCGTCACTCTGCATCTTCTAGTACAGTCAACCACTCACCATGTTTGCA	79146
Qy	1126	-----	1126
Db	79145	AAGCAGTCACACTGTCTCTTCTGTAGACATGGGCTTGGGCTGTGTGTGGCCTGGCCACA	79086
Qy	1126	-----	1126
Db	79085	GTCTCCAGTCGAGGGCCGCGAGCCGCTCTCTCACTTGGGTATCTTAAAGTCCTAAGCA	79026
Qy	1126	-----	1126
Db	79025	AGGCTCTTCTGTGTCTGAATCTGGTCTTCCCGTGGCTGAGTCTGGTGTGATCTATCTC	78966
Qy	1127	-----GlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAl	1143
Db	78965	TTGCGCCCTAGCAGGAATGTCAGTGCAGCTTCGGTGCCTCCCACTCAGAGCATGCCAAGGC	78906
Qy	1143	aProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLys	1160
Db	78905	CCCCATGGGCCCTCTCACCATGGGACTGCCTCTCGCTGTGACCCCTAAGAAGCTGGGTGA	78846
Qy	1160	-----	1160
Db	78845	GTGCCCATGACCTTTTGGTTCTTATGCCCCGGGGGCGAGTCTTGGCCCTGGGAACTTAGAG	78786
Qy	1161	-----LeuAlaProPheSerGlyV	1167
Db	78785	GGGGGGCAGCTGAGACCTCTGTGTACCTCTCGCCCACTAGCCCTTCACTGAGGAG	78726
Qy	1167	allYsGlnGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValP	1187
Db	78725	TGAAGCAGGAGCATGTTGTCCCTCGGGTCTCAGGCTGGCCACCTGAAAGTCTGGAGTGC	78666
Qy	1187	roThrAlaGlnGluAlaSerValLeuArg	1196
Db	78665	CTACTGCTCAGGAGACCTCTGTGCTAAGAGGTAAAGTCTGGCCCACTCACCTTCAGTCTTA	78606
Qy	1196	-----	1196
Db	78605	GAGAGCTGAAGCCCACTTGACAGTGTCTCTTACCCTTGACTCCACAGGAGCCGCT	78546
Qy	1197	-----GlyThrAlaLeuG	1201
Db	78545	AGGTGGCTTTTAGACTTGGTCTAAGAGCTGTGCTTATGCTTCTTAGGGACGACACTGG	78486
Qy	1201	lySerValProGlyGlySerIleThrlyGlyIleProSerThrArgVal	1217
Db	78485	GCTCCACTGCCAGTGAAGATCATCCAAGGGCTTCCCCAGTACCCGGACTGCAGATGGCC	78426
Qy	1217	-----	1217
Db	78425	CCAGTACCAGGCTCTATCACCCACGTAGTGTCTGGGGCAGGGTGGGATGGCAGGCA	78366
Qy	1217	-----	1217
Db	78365	TCCTAGGAGGTTGGGATGGATGAGAGGACCTCCGGGGAGACGAGAGATCTTTGAGCTGA	78306
Qy	1217	-----	1217
Db	78305	ATGGTTGGTCACTCGGTTGGGTCAGGCTCAGGCTGATAGGATGAAGAAACAGGTGTCTATGT	78246

Qy	1217	-----	1217
Db	78245	GTCTCCAGCAGTTGGGTGTGGCAGAGACCAACACAGAAAGGATAGCCAGGCCCTTAACC	78186
Qy	1217	-----	1217
Db	78185	AGACTGTTCAGGAGCCATAGATCATAAAGTGTCTGGACTGAGAAAGTTACCAGGTGACCCC	78126
Qy	1217	-----	1217
Db	78125	AGTGACATGTTGCCATGGCCAGAAAGACAGAGGGGACAGAGGTGGCAGCTGCGCTCTTGA	78066
Qy	1217	-----	1217
Db	78065	CAAAATCACCAGTCGGTAGGCATGAGCCAGTGGTGGGTACCAAGGCTTAGTACCCACGGTGT	78006
Qy	1217	-----	1217
Db	78005	GTGGCCAAGCATACAGCCACAGCGAACTTCCAGTCACGGACCAACCCACAGAGGTCTTT	77946
Qy	1217	-----	1217
Db	77945	TTGACTCACTGTAGGGGACAAATGTCTCATTTGCCCCAGGCTTCCCTCGAAGTCTCTGG	77886
Qy	1217	-----	1217
Db	77885	AAATCTTGGGCTTTCAAAGGCAAGCTCAGAAGCTGGAGGGGAAATTCAGTCATGTCTCT	77826
Qy	1217	-----	1217
Db	77825	TCTGGCAGTCCAGTAGCCCTCCTCGACTCAGTCTTCCCTGTGCAGAAATGTTCTTTCC	77766
Qy	1217	-----	1217
Db	77765	AGGATAGGTACTCCAAGGAGCTCTTAGAGTGCCAGGTTTGACACAGAGCTTCCAGGGC	77706
Qy	1217	-----	1217
Db	77705	TCAGGATGCCGCTCACAGGCTGAGNAATCCCTTCCCTGGTTAGGGCTTGTGGAGAGT	77646
Qy	1217	-----	1217
Db	77645	GAGGTCAAGCCTCTGGGGCTATCTGTGGGATGATGACCCAGGCATTGATGTGACACG	77586
Qy	1217	-----	1217
Db	77585	GCTAACAGGTGGAGACCAACATCCCATCTGTGCTGACCGCCCTGGAAACCAAGGCCAC	77526
Qy	1217	-----	1217
Db	77525	AGAGCCCATGTCCCACTCAGCTCCAGGATTACACTCAGGCCCTTGGAGCGGGCTGCGGG	77466
Qy	1217	-----	1217
Db	77465	TGGGCCCCCTGTGCTCCTGTATTTCTTGTGTTTGTCCAGTGAGTTATACGGAGCAGAAC	77406
Qy	1217	-----	1217
Db	77405	CATGGCCTGGCTGTATTTTCGGTTGTGAGAGGCCCTCTAAGCTTGGAGACTCAAAACATC	77346
Qy	1217	-----	1217
Db	77345	AATACCCCCCTTTTCATCTGGAGCCTGCCTGACCTTTTCTACCGTATATACCTGGCCTAGCAA	77286
Qy	1217	-----	1217
Db	77285	GCAGTACCAGCCACTTGAGAGGCATTTCCCGACACACCCAGAAATCACCCCGCATGCC	77226
Qy	1217	-----	1217
Db	77225	CCACGGGGACCAACCAACAGAACAGTCCCAACGGGAGACAAACATCACTGAATCTGTCTCC	77166

Db 75005 CCGACACGCTGAGGCTTCAACATGTCCTTGGACCCGGTGTCTACACATCCCTTGTGTACA 74946
Qy 1281 ----- 1281
Db 74945 TGCAGACCTGCAGCGCTGTGGGGCTGTCTCGGGACTCAGACCGTGTGACATCTTC 74886
Qy 1281 ----- 1281
Db 74885 GTTAGTGTGCACATACACCCATGGGCACAGACACATACACCCCTTACACAGCAGTGG 74826
Qy 1281 ----- 1281
Db 74825 CAGGCATGCCCTGGCAGCTTTCTGCCAAGCAGAGTTCGCCAGTTGAGCTGGTTGAGGGTCCC 74766
Qy 1281 ----- 1281
Db 74765 ATCTGTGGGGCAGGGCTCTCTTTCAGCTCTGCTGTTTGCCTAATCCATGCCCTGTGGT 74706
Qy 1282 ----- GlyGlyMetSer 1285
Db 74705 GTCTGCACCTTGTCTGCCACCCACACCATGTGCTTTTCTGGCCCTCAGGTGTATGTGG 74646
Qy 1286 ValThrGlnCysSerIysGluAspGlyArgSerSerGlyProProHisGluThrAla 1305
Db 74645 GTGTCACAGTGTCTCAAGGAGGATGGGAGGAGCTCAGGCCACCCACGAGACTGCC 74586
Qy 1306 AlaProIysArgThrTyAspMetMetGluGlyArgValGlyArgAlaIleSerSerAla 1325
Db 74585 GCCCTAAACGCACCTATGACATGATGGAGGGCGTGTGGCAGGACCATCACCTCAGCC 74526
Qy 1326 SerIleGluGlyLeu ----- 1330
Db 74525 AGCATCGAGGTAA-GTGTCTCACTGTAGTAGTTACCGGGGTCAAGCACCATCCAGGA 74467
Qy 1330 ----- 1330
Db 74466 GACCCAGGTACCAGGTAAAGATGGAAGGGAGGTGGCCCTTGATCATGCTGGAGGG 74407
Qy 1330 ----- 1330
Db 74406 AGCCAGTCAGTAGTTGGCTCAATATTAAACAGTAGAAGAACTGAGGCACACCCGGTGCA 74347
Qy 1330 ----- 1330
Db 74346 AAATGCAGCCCGGAAACGGGCACCTTGTCTGGGTGTTAAGTCACAACAGAGACTG 74287
Qy 1330 ----- 1330
Db 74286 TTGGTGTAGAGCCAGCCCTCTGGACACCATACGGTGAACATATCCCTGCTCTCCCA 74227
Qy 1331 ----- MetGlyArgAlaIleProProGluArgHisSerProHisGluLeuGluGln 1348
Db 74226 GGACTCATGGCCGTGCATC-CTTAGCAGACACAGCCCC-CACTCAAGAGACAA 74173
Qy 1349 HisIleArgGlySerIleThrGln ----- 1357
Db 74172 CATCATCCGAGGCTCCATCAGCAAGGTATATCTCTTACGGTGTGTAATTCACCT 74113
Qy 1357 ----- 1357
Db 74112 CCAGCCAAAACACAGCTAAGGGGTGGAACCCGAGGGTGAACCTTCCCGTCCGTCTTAT 74053
Qy 1357 ----- 1357
Db 74052 AGTGGAGTACTACTCTGTACTTGGCGCTCATTTCCAAAGCCCATGTGTCATGCCCTTGTCT 73993
Qy 1357 ----- 1357
Db 73992 AGTGATTGAGAGGGGTCCAGATTGAATGTCTTTTAGCTCATGCCCTTGTGTTAAACTG 73933
Qy 1357 ----- 1357
Db 73932 AGCAATGTGTTTCACTTAGCGTCTCTCAGTCTCATATGAAGAAACAGGCTTAGAAAAAGTCA 73873

Qy 1357 ----- 1357
Db 73872 GTTTGCATACCAAGTCCACAGGAAGATTGTGCTGACCCAGGCTGAGCCAGCAAG 73813
Qy 1358 ----- GlyIleProArgSerTyValGluAlaG1 1367
Db 73812 GCAGTGGAGAGCCAAATGATGTTTCCCGGAGGATCCCGAGGTCATATGTGGAGGCCA 73753
Qy 1367 nGluAspTyLeuArgArgGluAlaLysLeuLeuArgGluGlyThrProProProPr 1387
Db 73752 GGAGGACTACTTACGTAGGAGGCCAAGCTTTTGAAGCGAGAGGACGCCACCGCCCC 73693
Qy 1387 oProProSerArgAspLeuThrGluAlaTyLysThrGln-----AlaLeuGlyPr 1404
Db 73692 ACCGCCACCTCGGACCTGACTGAGACCTACAAGACCCGCCCTCGACCCCTCTGGGTCC 73633
Qy 1404 oLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSe 1424
Db 73632 CCTGAAGCTGAAGCCAACTCATGAGGGTGTGGTAGCAGCTGTGAAGGAGGCGGCCGCTC 73573
Qy 1424 rIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProAr 1444
Db 73572 TATCCATGAGATCCACGAGAGAGCTGCGCCGACGCCGAGCTGCCCTTGGCACCACG 73513
Qy 1444 gProLeuLysGluGlySerIleThr----- 1452
Db 73512 GCCTCTGAAGGAGGGTTCATCAC-CCAGGTACGAGGAGCACAAAGGAGGCGCGGGCT 73454
Qy 1452 ----- 1452
Db 73453 TGGCAAGCCTGAGGGAGTTCTAATGTCGCCCTCATTTGCATTTGGGTTATAACTAAGGCC 73394
Qy 1452 ----- 1452
Db 73393 CAGAGAGTCAGCGGCTTGTCTTGGTCACTGACAGGCTTGTGAAACCAAAGTCTTTGG 73334
Qy 1452 ----- 1452
Db 73333 TCACCTTGGCTGCTCCCCACATTTGCTACCTGCTGCTGTGTGTCAAGACTAGAGCCACTGT 73274
Qy 1452 ----- 1452
Db 73273 CCCCTACCTTGGTTCATGGCTTGATGGAAGAAAGAAACAAACAGCCAGAACGACCTA 73214
Qy 1452 ----- 1452
Db 73213 GGGAAGGCTGTGTCTACCTGAGCCAGGAGACACACAGTGAATGGACCAGATGATCTAGG 73154
Qy 1452 ----- 1452
Db 73153 GTATTCTTTTGAAAACTTAGTGTGCTCTCTCATTTTGGGCCCAGCACTTGGGAGCGGAG 73094
Qy 1452 ----- 1452
Db 73093 TCAGGAAGATTGTAAATTTGAGGCTAGCTAGGCTACAAAGCTGTACTCTGGCTTACTGC 73034
Qy 1452 ----- 1452
Db 73033 CGATCAAAAGTTTACAAATTTAATGAACAACAACAATTGAGTGAAATTGAGATCTGG 72974
Qy 1452 ----- 1452
Db 72973 GACCTAGTGTCTCACTGTCTCCAGGCTGGATCCCTGTGGGATCCTCTCCAAGATGTC 72914
Qy 1452 ----- 1452
Db 72913 AATAAATTCGGTCTCCGTGGTGTCTCTGAGAGCCAAACAGGAAATGTGTAACTCCAGCCT 72854
Qy 1452 ----- 1452
Db 72853 GGGACCCGGAGCCAGAGCTGGGTGACAGCAGATGCATCTGTATATGAGTGCATAGTGTCT 72794

QY 1452 ----- 1452
Db 72793 GCAGGATGTCGGGGGCCACACTGGGTGAGGAAGGTGCAATTGACCTGTGTAGTACCGACGC 72734
QY 1452 ----- 1452
Db 72733 TCTCGTACCCTCCACAGTGGGGGTCTCTGGACGTATAGCCAGGCTGAAGCGAGCAGG 72674
QY 1452 ----- 1452
Db 72673 TATCTAGCCTCTGACGTCTGTCTTGTCTCAAGGGGTTAGAGATGCCCATCTCTGTGCAGAC 72614
QY 1453 ----- GlnGlyTh 1455
Db 72613 ATACAGGAGAACAGCATCTGGCCGCCACCATACAGACTGGACCTCTCCCCAGGGGAC 72554
QY 1455 rProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSe 1475
Db 72553 CCACCTCAAGTAGACTCCGGGGTCCCTCCAGTGGCTCCAGAAACACGACGTACGCTC 72494
QY 1475 rLeuileGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAs 1495
Db 72493 CATCATCGCAGCCCCGGCGGCTTTTCTGCGCTGCACCCACTGGACATAATGGCTGA 72434
QY 1495 pAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAl 1515
Db 72433 TGCCCCGGCACTAGAGCGCGCTGTCTATGAGAGAGTCTGAAGAGCGGTTCAGGGACCG 72374
QY 1515 aserSerSerGlyLysSerileAlaArgGlyAlaProValileValProGluLeuGlyLy 1535
Db 72373 CAGTGGTGACGGGGCTCCATCACACGCGGGCCCCAGTTGTGCTGCTGAGCTAGGCAA 72314
QY 1535 sProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuPr 1555
Db 72313 GCCCCGGAAGGCCACTGGCTACGAGGACCAAGGGGCGCCCTTACCAGCCACCTGCCC 72254
QY 1555 oArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSe 1575
Db 72253 CCGAGGCTCACTGTGACTACAGGGAGCCACGCCACGGCTTCAGGAAGGTGAGGC-TC 72195
QY 1575 r----- 1575
Db 72194 CCCTGGGTGCCAGCTGTGGGGGGGACAGCTCATTTTCCACGGCAGTCTGTGCTG 72135
QY 1575 ----- 1575
Db 72134 CCCTGCTCCACACTGCATACAGAGCCCCCGTTCCTCCCTTACCAGCATGA 72075
QY 1575 ----- 1575
Db 72074 TAATCAATCAGGGCTGAGGAGGGCTCAGTAGCTAAAGTCTTGTGTACCAGCATGA 72015
QY 1575 ----- 1575
Db 72014 GACTCAGAGTTCAAATCCCCGGTACCCACATAAGCCAGGCACACTCATACACTCTGTAA 71955
QY 1575 ----- 1575
Db 71954 TTCTCATGCTTTCACAGTGAGACAAAGGCTAAGACAGAAATCTCTCTAAAGGTTATAGGCCA 71895
QY 1575 ----- 1575
Db 71894 TGAGTCCGGGTCTCAAAATGCGGAAGTGGATGACACCTCTCCCCCCCCCTCTATT 71835
QY 1575 ----- 1575
Db 71834 TTATACACGCATACGCACACTCATATCAGACACACTCGCGCTACATTCTCTGTGCAC 71775
QY 1575 ----- 1575
Db 71774 ACATGTACACAAAAACAATAGTTTAAATGAAGAGTTCCCACTCTGTAGGCTTCACCT 71715
QY 1575 ----- 1575

Db 71714 CTCCAGTAGCGTCTGCTGTGAAGCACCTGCTGTGACGTGACCGCGCATCTCTACTCT 71655
QY 1575 ----- 1575
Db 71654 GAATCCGCTTCTGATTTCTTCGCATTTCTGACCTCCGCTCTTTCTCCGATCTCCACAT 71595
QY 1575 ----- 1575
Db 71594 GTGAGCCTTTCCGTGTGTGCTGCGCTCTATGCCGAAGGCTCCACCTTTGGTTTGTCTCT 71535
QY 1575 ----- 1575
Db 71534 AGCCAAACCACTGTATATACCCACGTCAACAGATGTCAAGGCGCTGGAGAGCTATAGGCAC 71475
QY 1575 ----- 1575
Db 71474 TCCAACTAGTTCCAGCCACATCCGGCGCGGATGTGTTAGGTTGTACAGCAGTGTCTCT 71415
QY 1575 ----- 1575
Db 71414 AATTCAATTGATGAGGAGGCCCTTCCAGAGCACTGAGATCTCAAACATCTCTCTCTCT 71355
QY 1575 ----- 1575
Db 71354 CAGGCCGTCACCTTCTACCTTTGGGTTTCATCTCTATCTGGAACAATGGCTCTTGGAG 71295
QY 1575 ----- 1575
Db 71294 CTCAGGGTTCTTGGCTGTCTTCTCTCAGGGTGTCCCTTCTCTCTCATGGAACTGCC 71235
QY 1575 ----- 1575
Db 71234 AGCCCATGATTTAAGTCAGACTTGAATGAATAGGCAATGGTGACTGGGAGTAAATGAA 71175
QY 1575 ----- 1575
Db 71174 TGCTTTCAAGAGGAGCAGGAGCAGAGACAAGAGGCTTAGGGCTTGTGACAGCTCAG 71115
QY 1576 ----- -Se 1576
Db 71114 CTGTGGCCCCGTTGAGGCTGAGGCCACTTTGACCTGTCTCCACAGGCGCTCTCTC 71055
QY 1576 rSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerPr 1596
Db 71054 CAGCAAGGCTTCCAGGACCGGAAAGCTGACCTCGACACCCCGGAGATTGCCAAGTCCCC 70995
QY 1596 oHisSerThrValProGluHisHisProHisProHisProHisProHisProHisProHis 1616
Db 70994 ACATAGCACTGTGCGGAGCAGCCACCCCTCCCTCTCCCTCTAGAGCACTGTCTCCG 70935
QY 1616 gGlyValSerGlyValAspLeuTyrArgSerHisHisProLeuAlaPheAspProThrSe 1636
Db 70934 GGGTGTGACTGTGTGTGAGCTGTACCTGTGGCCACATCCCATTTGGCTTTGACCCCACTC 70875
QY 1636 rIleProArgGlyIleProLeuAsp----- 1644
Db 70874 CATACCCGAGGATCCCTCTGGAAGCAGGTGGGTGCTGTGGGCTGTGTACATGTGGGGCTT 70815
QY 1644 ----- 1644
Db 70814 ATTGATGTGCTTCTTGGGTGGACACTTGGGGTATGGCTTAGTGTGCTGCCATCACCAGCT 70755
QY 1644 ----- 1644
Db 70754 GCTTACCCAGGCGAGTTCTTAGGATTTGGGCTTTTGTGGTGTCTTCTACCGTAGGACT 70695
QY 1644 ----- 1644
Db 70694 CATCTGCTCTGCAAGTGACGATAAATAGGGGCTGCAGTGAGTGCCATGATGTGGAACC 70635
QY 1644 ----- 1644

Db	70634	TTTTAAAGGCTCCTTTGTGAAGGTGTAGACCTGTATTTCGATGCAATTATATGGGG	70575
Qy	1644	-----	1644
Db	70574	GCACCTCCGAAACATTAAAGACAGGCCCCCACATGATGAGGGAGCGTCTTAAGCTC	70515
Qy	1644	-----	1644
Db	70514	TGTGTGTGCAATTGGGGAGGCATGTGGGTGTACAAGGAGAGGAAGCTGTGTGGCAGAAGT	70455
Qy	1645	-----	1645
Db	70454	GAGGAAGCTGAAGATAAGTCATGGGCCACCTGACCTCCATCCCTCCCTCCCCAGC	70395
Qy	1645	-----	1645
Db	1645	aAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTy	1665
Db	70394	AGCGCAGCCTACTACCTGCCCCCGGCACCTTGGCCCCCAGCCCCACCTACCCACACCTGTA	70335
Qy	1665	rProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIl	1685
Db	70334	CCGCGCTTACCTCATCCGCGGCTACCTGACACCGCGGCGCTGGAGAACCGCCAGACCAT	70275
Qy	1685	eIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGl	1705
Db	70274	CATCAATGACTACATCACTCGCAGCAGATGCACCAACACGCGCTCTGCCATGGCCCA	70215
Qy	1705	nArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAl	1725
Db	70214	GCGGCGCATATGCTGAGGGGTCTGTCCCGAGAGAGTCTCTCGCTGSCCTCAATTATGC	70155
Qy	1725	aAlaGlyProArg-----	1729
Db	70154	CGCTGGCCCGCTGGTGGTTCAGTTGCTGGGGGTGTGCCCATCAGCAGAGACTTGTCTTCGCG	70095
Qy	1729	-----	1729
Db	70094	TGTGTTGTGCAAAAGAAAGTGCACACGTGTGCATCGCACTGGGCGGAGGGGCACAGTT	70035
Qy	1729	-----	1729
Db	70034	GGCTGGGAGGACTAGGTGGTGCATGTGGCTCTTGGGGGAAGGGCAGAGGGGCCACTT	69975
Qy	1729	-----	1729
Db	69974	GTATCTGTTCACTATAGAACTGTGCTTAGCAGACACTTTGCCACCATGCACGTTGGGT	69915
Qy	1729	-----	1729
Db	69914	CTTTGGGTGTGTTCTGAAACCTTGGTATGGGGCATCTTCTGTATCCAAAGAGATCAAAA	69855
Qy	1729	-----	1729
Db	69854	TACAGGCTGATGGCAATTGCACGCCATATCTTAAGTAGAGGAAGGTGCAGGCCAG	69795
Qy	1729	-----	1729
Db	69794	CCTCAGCTACATAAAATTTTCAGGCGACCTGGGCTATATGAGTCTGCTCTATACAAATG	69735
Qy	1729	-----	1729
Db	69734	ACTGAAAGAAAGGAATGAGAGCATGCGGCTTGTGCAGAACACTGACTTCTGTGAACAGC	69675
Qy	1729	-----	1729
Db	69674	CTGGGTCACTTTGCTCGACTGTTTCAGAGATAACCTTGTGGCCCCCAGTTAGTATACGGA	69615
Qy	1729	-----	1729
Db	69614	AGCCACTCTGGGGTTTTAGAGATGGGAGCCCCAAATGCAAGAGAGTGTACTGTGGGAAG	69555
Qy	1729	-----	1729
Db	69554	GCAGTGTGTTGGGTCAAAAGATGTTGTTTGAGAGACACTGCACGTCTACCTTGAGAAA	69495
Qy	1729	-----	1729
Db	69494	GGGGCTGGGGCCACTCTCTGGCACCCAGGAGGGTAGTCAGGTGAAGGTCTCAGGCCT	69435
Qy	1730	-----	1730
Db	69434	GCACCTTGGGGACAGACATCAGGCACCTACAGCTGACATCCACCCACCCTTCGAC	69375
Qy	1730	-----	1730
Db	69374	GCATTATCGACCTGTCCCAAGTGCACACTGCCCGTCTGTGGTCCGCCCAACTCCAGCA	69315
Qy	1750	hrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerS	1770
Db	69314	CCCTGCCACCGCATCAGCCGCTTGCTTACCTCCCACTGGGCCCCCACTTCAGCA	69255
Qy	1770	erArgHisSerSerProLeuSer-----	1778
Db	69254	GCCTGCACAGCAGTTTCACTTGTCCCACTGTCCCACTGGCTGCCAAATGGCTGGTGGGTCT	69195
Qy	1778	-----	1778
Db	69194	GTGCCTCACCCCTGTGCTCCTGCAGTAAGTTCTGGCCCACTGGGCTCTGGCTCACACTC	69135
Qy	1779	-----	1779
Db	69134	TGGCTTGGGGTTTCCAGGAGGCCCACTCACCTAGCTAAACCACTGCCACGTCATCAT	69075
Qy	1794	erGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerI	1814
Db	69074	CGAGCGGGAACGGGAACGAGAGCGAGACGAGAC-----	69033
Qy	1814	leLeuThrSerThrThrValGluHisAlaProIleTyrArgProGly-----	1830
Db	69032	TCCTCAGCTCACCACTTACAGTGGAGCATGCCATCTCGAGACCTGGTAGGSCACCCC	68973
Qy	1831	-----	1831
Db	68972	AGACCCGCGCCACCCAGGCTCTCATGGGCCACCTGGGCTGCCCCCTGACTCGCGG	68913
Qy	1832	luGlnSerSerGlySerSerGlyGlyGlyGlyGlyGlySerSerSerArgProA	1852
Db	68912	CTCG-TCAACAGGTACGAGCAGCAGCGGGCTGGGGC-----	68860
Qy	1852	laSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuG	1872
Db	68859	CTCCCCACACC-----	68806
Qy	1872	lnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluP	1892
Db	68805	AGCAGAGGCCCACTGTGCTGCACACACGAGCATGAAGGGCGTGGTCACTCCGTGGAAC	68746
Qy	1892	roSerLysProThrValLeu-----	1898
Db	68745	CCGGCACGCCCACTGCTGAGGTGGGCCAGGTTGGCACCAGGGGGAAACAAGCCTGGGG	68686
Qy	1898	-----	1898
Db	68685	GAGGGTACACAGATGGAATTGCTGAGAGGGAGGGGTGGGGGTGGGCAGATGGGTAGGG	68626
Qy	1898	-----	1898
Db	68625	AATGGAGCAGATGCGAGGATCGGAAGTGGGAAGTTCAGAGAGACTATGAGATGGGGAT	68566
Qy	1898	-----	1898
Db	68565	GGGTGGCTTGTACGCGGGGGAGGATGGCCCGAGAGAGGGGCACACAGTGAAGTGGACGG	68506
Qy	1898	-----	1898
Db	68505	GAGATAGTGCAGATAACAGGCAGATGACAGTGGGTGGATGAAGATGATGATGATGATG	68446

Db	66297	ACCACCTGCACCTTCGTTTTTGGGAGTGGGATGACGGGTCTACAGTTGACTGTGCCAC	66238
Qy	2039	-----	2039
Db	66237	CACGTACATGGAGAGTTAAGCAGGAGTGGTGGGGGAGGCTGGTACCCACC	66178
Qy	2039	-----	2039
Db	66177	TCCTCTCCACACACAGCTACAGCCCTAAGGGAATGAAGCTCCCGCTGGATCAG	66118
Qy	2039	-----	2039
Db	66117	AGACAGTGGGACTGGGCCAAGCTGGACCTAGGGACCCATTGTCACTCAGCTTGAGCCT	66058
Qy	2039	-----	2039
Db	66057	CACACACACACACCTGACCCCATGACATTCACACCTCCCTCCAGTTTCTCCACACC	65998
Qy	2039	-----	2039
Db	65997	ACCTCCAGTCTCCCTATCTGAGCCCCACACGCCCAAGTCAGACTCCCATATCTTT	65938
Qy	2039	-----	2039
Db	65937	GGGTGTCTGAGTCTCTATAAATGGAGCTCCCTTCACTTGCAGAGCTTCTCTTGCC	65878
Qy	2039	-----	2039
Db	65877	ACCCCTAGTTCGACCTGATTTTCTTCCATCTTGAGCCCGACAGTGGAGGAATGGCTCC	65818
Qy	2040	-----SerLeuGlyTyrHis---GlyS 2046	
Db	65817	TCCCAAGCCCCACCATCTTGACCCCTAGCTACCTGCTCTCA-GGTTACACAGTGGAG	65759
Qy	2046	erSerTyrSerProGluGlyValGluProValSerProValSerSerProSerLeuThrH	2066
Db	65758	CTGGCTACAGCCCTGTAGGGGTAGAGCCATCAGCCCGTGAGCTCGCCAGCGTGAACC	65699
Qy	2066	isAplysGlyLeuProLysHisLeuGluGlyLeuAplysSerHisLeuGlyGluL	2086
Db	65698	ACGACAAAGGGGCTCTCCAGCCTCTGAAGAGCTAGAGAGAGCCACTTGGAGGGGAGC	65639
Qy	2086	euArgProLysGlnProGly-----	2092
Db	65638	TGGGCACAGACGACGACGAGTGCACCCGGGAGAGTACAGGTGGGAAACTGAGGCACGG	65579
Qy	2093	-----ProValLysL 2096	
Db	65578	CTCCATGTTCTCCTCAGGCCCTGACAAAGCTCTGTTGTCTCCTCAGGCCCATG-AAAGC	65520
Qy	2096	euGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerS	2116
Db	65519	TCAGTGGGAGGCTGCCCATCTCCCATCTGGGGCCACTGGCCGAGAGCGAGCCCTCAT	65460
Qy	2116	erSerProLeuLeuThrAlaProGlyValLysGlyHisGlnArgValValThrLeuA	2136
Db	65459	CCAGCCCACTCTCCAGACTGCCCCAGCATCAAGGTCAAGAGTCCACAGAGGGTGTACCTTG	65400
Qy	2136	laGlnHisIleSer-----	2140
Db	65399	CTCAGCACATCAGTGAAGTGCCCATCTCTGGCCCTGCTTGGGAAGGAGCGTGGCTA	65340
Qy	2140	-----	2140
Db	65339	TGACAGCCATATCCACAGCCACAGACAGATGCTAATGCTTAATGTCGCCACCGCTTAG	65280
Qy	2140	-----	2140
Db	65279	TCCTTTGGGGGTTCATCATAGGTCTAACAGGACACTTACCCAGGAGACCAAGAGCCAG	65220
Qy	2141	-----GluVal 2142	
Db	65219	ATCAGGGTCTGCGCTTACTCTGAAAGTCTGAAAGTCTGAAAGTCTGCTTCTCCTGAGGAGGT	65160
Qy	2143	IleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaPro	2162
Db	65159	ATTACACAGACTACACACGCCACCCACAGAGCTCAGTGGCCCCCTTCCCGCCCT	65100
Qy	2163	LeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAsp	2182
Db	65099	CTCTACTCTCTCCGTGGAGCAGCTGCCCTGTCTGGATCTCCGCCCTCCACCCAGTGAC	65040
Qy	2183	LeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGly	2202
Db	65039	CTCTACTCTCCACCCCGGACCATGGCACCCAGCCCGGGATCCCCCAGTGAAGGG	64980
Qy	2203	GlyLys-----	2204
Db	64979	GGCAAGAGGTGAGGAAGTACCCATGTCTGCATCAGGTGCTGGGAACAGTGGGACAGGGGT	64920
Qy	2204	-----	2204
Db	64919	GAGCTTTGAGTGTGGGCTGATGGGAACTGAGGTGAGGGATACAAAGCTCCTCAGTACG	64860
Qy	2204	-----	2204
Db	64859	CTCCAGAACAGACACAAAGCATCTAATCAAGAAGTGGCTCACCCAAAGGTCCACGTG	64800
Qy	2204	-----	2204
Db	64799	CCAAATTGAATGTCCTATATATACCACTCAGTAGGGGTGAGTGTCCCATTTCAATT	64740
Qy	2204	-----	2204
Db	64739	GCAGATAGGAAACTGAGGCATAGAGAGGTGAGTTACTATCCCGTCCCATGCACAAAC	64680
Qy	2204	-----	2204
Db	64679	ACCCTGCACAGATGTAGACAGGTATCTCTGGGTAAACAGCGAGTCTTAGCATNGGGGCT	64620
Qy	2204	-----	2204
Db	64619	CATCAGGGCTTACAAATGCTCCAGCACCAGCAGCACTGTGTGGTGGCATACCTC	64560
Qy	2204	-----	2204
Db	64559	TGCACAGGCTGAGGCAAGACGACTGTGAATGTCAGGCGCAGCTGGGTGTACCAAGAC	64500
Qy	2204	-----	2204
Db	64499	CCTGACTCAAAATAAAACCAAGCCCATGAGTGTGGGTGAGTTCGAAAGTCTCAGAAG	64440
Qy	2204	-----	2204
Db	64439	CCCCCTGGGTCTCCATACCTCAGATGCAGAGTCTGCCATTACAGGGAAACACGTGAG	64380
Qy	2204	-----	2204
Db	64379	CCCCACTGTGGCAGTAGGGGTCTCTTCATGAGTAGAAACCCCACTCTCAGCTGGAT	64320
Qy	2205	-----ArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAsp 2220	
Db	64319	GCTGTCTCCATTAGGTCTCCAGAACCCAGCAACA---GTCCTGAGTAGCAGCGAGAT	64263
Qy	2221	GlyIleGluProValSerProGluGlyMetThrGluProGlyHisSerArgSerAla	2240
Db	64262	GCTATTGAACCTGTGTCCCGAGGGGATGACTGAGCGAGGATGACGAGGACTGACGAGGACT	64203
Qy	2241	ValTyrProLeuLeuTyrArgAspGlyGluThrGlu-----	2253
Db	64202	GCGTACCCACTGTGTACCGGACGCGGAGCAGGGGAGCNCAGGTAGTAGTATGGCACA	64143
Qy	2253	-----	2253
Db	64142	TCCCCGTTAGAGGTGGTGGTGTGGTTTCCCATACACCCGTGGCTTCAGCCCTCTAGCG	64083

Db	61923	AGTCATGTTTCCAGGCATGGCCAAATGACACTCAACATCGGGCCAGAGGTAATGAGGAAG	61864
Qy	2322	-----	2322
Db	61863	GTCAATGCTGTGCCAGTGTAGCGCCCTCAGGTACAAAGCCAGGAGGGTCACTGTCTCC	61804
Qy	2322	-----	2322
Db	61803	ATCAGAAACTCCCTCCACCCCTCTTCTCCTGTTACCTCGTATCTCCCTCTGTATT	61744
Qy	2322	-----	2322
Db	61743	CTGGATGTGAGAGAGTGGGTGGGCTTGTATTATTGAGGTTAGGGCGTGAGGGGTGTGT	61684
Qy	2322	-----	2322
Db	61683	GTGTGCGTGTGCGTGTGCGTGTGCTGTGGCATGCAATTTACCCCTGTTCAGAGTTTCGGCT	61624
Qy	2322	-----	2322
Db	61623	TTTTCTTGGTAACATTTGCTTAGCACTGAGGGAACGTGAGCATCCCTGACCCGAGCTTGC	61564
Qy	2322	-----	2322
Db	61563	CCGAAACCCAACTGAGGCTCCACAGCACCCCCAGAACAGGATTAGGAAGTACTGTCCC	61504
Qy	2322	-----	2322
Db	61503	TTCCATCTCAGAACCTGGCTATTTTACATTTGCCCGCATTTTACGCCATGTGATTTTCAC	61444
Qy	2322	-----	2322
Db	61443	AAGTTTTGGTAGCCTGGAGGTTTACGGGGACCAGAAACTGGCCAGTGGCATAGGCTGGG	61384
Qy	2322	-----	2322
Db	61383	AATCCTGCCGATGACGTGTGTCTGTGTTTTTGTGTGTGTGTGATTTGGGGTTTTTA	61324
Qy	2322	-----	2322
Db	61323	TTGTTGTGTTTGTTCATTGTTGTTGTTTGGGGGTGTTGATGAGTTTAGCTTG	61264
Qy	2322	-----	2322
Db	61263	GCACTGTTTGTCTTGTGGGCTGGGTTTGGAGCAGTTTTTGTGGGTACTTGGGATTGAGG	61204
Qy	2322	-----	2322
Db	61203	CAAGACCTTTGCACTCACTAAGCAAGTGTCTCCTGTGTGTGAGCTACAGCCAGCCCTC	61144
Qy	2322	-----	2322
Db	61143	CTCTGCCGAGAGAGGTTCTCTAGCTGCCAGCCCGGCACAGATTGAGCTGAACACTACT	61084
Qy	2322	-----	2322
Db	61083	GTAGACTAAGCTAGCCCCAACCTTGTCTATGCTCCTGCGTAGCTCATAAATAGCTGGGT	61024
Qy	2322	-----	2322
Db	61023	TTCCAGGCTCTTAATGAGGCTCCAGCCCTCGGGAGGTTCTGGAAGCATCCACTGAAAAAG	60964
Qy	2322	-----	2322
Db	60963	AGATGCTCCCCACCTGGCATTTTGCAAAAGTCCCAGAGCACAAATTCATGAGAGGAGAG	60904
Qy	2322	-----	2322
Db	60903	GAAAGAAACCGGAGAGGCTTTGAGAGGCTCAGTGCAGGCTGCAAACATCTCTTGAGGC	60844
Qy	2322	-----	2322
Db	60843	CAGTGTGTATTGTACTACTCTGTCTTGTCTGTCTGATGGGTTGGGCCAGCCAGGCTGGAGT	60784
Qy	2322	-----	2322
Qy	2322	-----	2322
Db	60783	GGCCCTTCACTGTATTAGCGCTACCTGTCTCTGCAAGCACATAACTGCATGGCTGTGAC	60724
Qy	2322	-----	2322
Db	60723	CCCCCATCCCTTGTATCCAGCTCTACCTTTCTCTGGCTCCTGTGGGGAATAATGGGGTG	60664
Qy	2322	-----	2322
Db	60663	GAGGGGGCGAGCTAGCTAAGGGGGAATCAGCAGGCTGAGGTCCCTAACCCCTGAGA	60604
Qy	2322	-----	2322
Db	60603	AAAGGGGAACAACAGGGGGTCTCTCTGTGTACCTTTTGTGTCACCATCCCTTCTCTAGAA	60544
Qy	2322	-----	2322
Db	60543	CCTACTCACCAGAAGACCCCGGCTGGACTTCTATCCCGTCCCTTAGCACCTAGCTCTGGC	60484
Qy	2322	-----	2322
Db	60483	TTAACTCCATAGCACAGCTGCCCCCTGCTTCTTTAGACATCCACCTTATAGTCAAAATG	60424
Qy	2322	-----	2322
Db	60423	GAATAATAGACCCAGCTGCGATTGGGGTGTAGAGTAGGAACAGGGCTGGGGACCTCCG	60364
Qy	2322	-----	2322
Db	60363	GCTGCCCGACCCCTCCCTGCAACAGCCATGAGCTGTGGGGCAGTTAGGGACCCCATC	60304
Qy	2322	-----	2322
Db	60303	CGCACCTTACACAGAGGGCGGGCTGGAGGACGAGATGCTGCGAGATGTGTCTCAC	60244
Qy	2322	-----	2322
Db	60243	TGTGTCTCTTCTCTCTCCTCCTCCTCATCTCTCGCAACCCCTCTGGCTCCCGG	60184
Qy	2323	-----GlyLeuMetThrTyrArgSerGI 2330	
Db	60183	CCCCCGCCCCCTGTGTGTCTGTCTCTCTCCAGGCTTATGACCTGTAGAAGCCA	60124
Qy	2330	nAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLe 2350	
Db	60123	GGCGGTCCAAAGAACACCCAGCACCAACATGGGCTAGAGGCCATTATTAGAAAGGCACT	60064
Qy	2350	uMetGlyLysTyrAspGlnTrpGluGluSerProLeuSerAlaAsnAlaPheAsnPr 2370	
Db	60063	CATGGGTAAATATGATCAGTGGGAGAGCCCCCGCTCGGCGCCAATGCTTTTAACCC	60004
Qy	2370	oLeuAsnAlaSerAlaSerLeuPro---AlaAlaMetProIleThrAlaAlaAspGlyAr 2389	
Db	60003	TCTGAATGCCAGTCCAGTCTGCCGCTGCGCTATGCCATACCACTGCTGCAGGACG	59944
Qy	2389	gSerAspHisThrLeuThrSerPro----- 2397	
Db	59943	GAGTGACCAACCACTACCTCGCCAGGTCTGCAGGCTGCCCCCTGCCACCCCGCCGCC	59884
Qy	2397	----- 2397	
Db	59883	GGCTGTGCTCCTTACCTGGTGTGTTTAACTCTTTCACCTCCCCCTCAAGCTTGG	59824
Qy	2397	----- 2397	
Db	59823	GAGACTGTGACCTTATTTTGTGTGCTTTTGACATCTAGTTTGTCTAATCCAAAGCTGTC	59764
Qy	2397	----- 2397	
Db	59763	TGACACCCCATGTTTGGAAAGGGGCTCCCCCTCGCCCCCAGCACGTGCTGTTCCTCCGTCG	59704

QY 2397 ----- 2397
Db 59703 TCACACCACTCCCAACATCCCAACCCCTTCTCCCTCTGCTGTGTGGACAGGATCAAGCTG 59644
QY 2397 ----- 2397
Db 59643 ACCAGGCCCAACCATGTGCGACCGGGGCTACTGCTGGGAAGAGTCTAGTGTGGGATGACC 59584
QY 2397 ----- 2397
Db 59583 CTTTGAAGGCCAGGCAGCAGCAGAGCCAACTCCCGGTTCTTCCAGTGCCTGTCAAGCACTCT 59524
QY 2397 ----- 2397
Db 59523 CTGTTCCCAAGCAACACAAAGCAAGAGGTGCTCTGCTGTCCCTTCTAGGGTCAGGCCA 59464
QY 2397 ----- 2397
Db 59463 GCCTCAGGGTCTGGCCCTATCAGCAGAGGCGCTTCAGTCTAGCACAAATTGAATGCTGTCTT 59404
QY 2397 ----- 2397
Db 59403 GGGCAATGCTGTCTCTCCAGCCAAAGATCTGACACTCTCCTAGTGAGGGGTACAGCCTGGTG 59344
QY 2397 ----- 2397
Db 59343 GGGCAGCTAAGGGAATAGTCAACCACTCTCTCATAGCAGATGTTTTTCCTCTTTTTC 59284
QY 2397 ----- 2397
Db 59283 ACGCTAGGGTGGCTCTGGCCATCTGTGTAGACCTCTTAGCCTCTGGCTCAGGGTTTCT 59224
QY 2397 ----- 2397
Db 59223 CACACTAGTTTCTGTGTTGGGCGCTAGGAGCAGGTGGCTCCCCAGGAGCAATCGGCC 59164
QY 2397 ----- 2397
Db 59163 CCACTGCCCTCTCCCTGCTGTGGTGACCCACAGTCAACACCATCTAACTCAGCC 59104
QY 2398 -----GlyGlyGlyGlyAlaLysValSerGlyArgProSe 2410
Db 59103 CTTTCCCACTTTGTCCTCGAGTGGTGGTGGGAAAGCCAAAGTTTCTGGCAGACCCAG 59044
QY 2410 rSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerVa 2430
Db 59043 CAGCCGAAAGCCAAAGTCCGAGCGCCAGCGCTGGCACCCGAGACCCACCACTTCTGT 58984
QY 2430 lSerSerValHiserGlyAspCysAsnArgArgThrProLeuThrAsnArgValTr 2450
Db 58983 CTCCTCAGTACACTCGAGGGGCGACTGCAATGCGCGAACCACCACTCACCACCCGTGTGTG 58924
QY 2450 pGluAspArgProSerSerAla----- 2457
Db 58923 GGAGGACCGGCCCTCATCTCGAGTGGGTATCAGTAGGGGCGAGAGTAGACATGGGG 58864
QY 2457 ----- 2457
Db 58863 TTCCTGGGTACCCAGGGGAGCGGCTCCCTGGTCCCACTAGCAAGGAAGGTGTAGG 58804
QY 2457 ----- 2457
Db 58803 AAGCCACACATTGAGAGGGTCTGTAGATCTCCTCAGGCGCTTCCAGTCAGGCTGGTGCAC 58744
QY 2457 ----- 2457
Db 58743 ACAGGGGACAGTGCCTGTCTCCAGCAGCTAGGATTTTGACAGGAAAGCAGGACGTAGA 58684
QY 2457 ----- 2457
Db 58683 AGGAAAGTTACCACAGCCGAGTCAAGGTGGTAGGCTGTCTAGCACAGGCAAGGG 58624
QY 2457 ----- 2457

Db 58623 AGAAGGCGCGCTTTTGAGAAACTCAAATGGTGTCTGGCTTAAGAGGCCCGAGCTCGAGCTTC 58564
QY 2457 ----- 2457
Db 58563 CTGTGTGTAGTAATGGTCTGTAGGTGGCTACTTTGAGAGCCCTCTCCACCATCTCTCCA 58504
QY 2458 -----GlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaG 2474
Db 58503 TCTTTCCACAGGGTTCAGCCATTCCCTACAACTTTTGATATGAGGCTACAGGCAG 58444
QY 2474 lyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyP 2494
Db 58443 GTGTCAATGGCTCCCAACCCCACTGGCTCCCGCAGGCGAGGCCCTAGTGGTC 58384
QY 2494 roHisAlaTrpAspGluProLysProLeuLeuCysSerGlnTyrGluThrIleuS 2514
Db 58383 CCACCATGCTGGGATGAGGAGCCCAAGCCACTGTGTGTTCACAGTACGAGACACTCT 58324
QY 2514 erAspSerGlu 2517
Db 58323 CGACAGCCGAG 58313
RESULT 17
AC121005 219339 bp DNA linear HTG 19-NOV-2002
LOCUS Rattus norvegicus clone CH230-322A11, WORKING DRAFT SEQUENCE.
DEFINITION AC121005
ACCESSION AC121005
VERSION AC121005.4 GI:25097679
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 219339)
Muzny,D.,Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Vollen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalo,K., Blair,J., Burch,P., Burrell,K., Calderon,E.,
Bryant,N., Buhay,C., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Cardenas,V., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Chacko,J., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Cleveland,C., Cockerill,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinu,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.R., Lebow,H., Levan,J., Lewis,D., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Louised,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapa,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwankweliemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,

Qy	1217	-----	1217
Db	168523	AGACTGTGAGGAGCATAGATCATAAAGTGTCTGACTGAGAAAGTTACAGAGTGACCCC	168582
Qy	1217	-----	1217
Db	168583	AGTGACATGTTGCCATGGCCAGAGAAAGACAGAGGGGACAGAGTGGGAGTGGCTCTTGA	168642
Qy	1217	-----	1217
Db	168643	CAAAATCACCAGTCGTAGGCATGAGCCAGTGGTGGGTACAGGCTTAGTACCCACGGTGT	168702
Qy	1217	-----	1217
Db	168703	GTGGCCAAGCATACAGCCAGCGAACTTCCAGTCAAGGACCCACCAAGGGTCTCT	168762
Qy	1217	-----	1217
Db	168763	TTGACTCACTGTAGGGGACAAATGTCCATTATGCGCGCAGGCTTCCCTGAAATCTCTGG	168822
Qy	1217	-----	1217
Db	168823	AAATCTTGGGCTTTCAAAGGCAAGCTCAGAAAGCTGGAGGCGGAAATTCAGTCAATGTCCT	168882
Qy	1217	-----	1217
Db	168883	TCTGGCAGTGCCAGTAGCCCTCCCTGACTCAGTTTCCCTGTGACAGAAATGTTCTTCC	168942
Qy	1217	-----	1217
Db	168943	AGGATAGGTACTCAAGAGGCTTTAGAGTCCAGGTTTGACACAGAGTTCGACGGGC	169002
Qy	1217	-----	1217
Db	169003	TCAGGATGCCGGCTCACAGGCTGAGAAATCCCTTCCCTGTTAGGCTTGTGGAGAGT	169062
Qy	1217	-----	1217
Db	169063	GAGGTCAGCGCTCTGGGSCCTATCTGTGGGATGATGACCCAGGCATTGATGTGACAG	169122
Qy	1217	-----	1217
Db	169123	GCTAACAGGTGGAGACACACATCCCACTGTTGCTGACCGCCCTGGAACCAAGGCCAC	169182
Qy	1217	-----	1217
Db	169183	AGAGCCCATGTCCTCACTCAGCTCCAGGATTACACTCAGGCTTGGAGCGGCTGCGG	169242
Qy	1217	-----	1217
Db	169243	TGGGCCCTGTGGCTCCCTGTATTCTTTGTTTGTCCAGTGAGTTATACGGACAGAAC	169302
Qy	1217	-----	1217
Db	169303	CATGGCTGGCTGTATTTCGGTTGTGAGGCCCCCTCTAAGCTTGGAGACTCAAAACATC	169362
Qy	1217	-----	1217
Db	169363	AATACCCCTTTTCATCTGGAGCCTGCTGACCTTCTCACCGTATACCTGGCCTAGCAA	169422
Qy	1217	-----	1217
Db	169423	GCAGTACCAGCCACTTGAGAAGGCAATTTCCGACACACCCAGAAATCACCCCGATGCC	169482
Qy	1217	-----	1217
Db	169483	CCACGGGACCCACCAACAGAAAGTCCCAAGGGAGACAAACATCACTGAATCTGTTC	169542
Qy	1217	-----	1217
Db	169543	TGAGGTCATAAATGAAGAGAGAACTCAGGATTTGGGGCCCTCTGAGAGCCAAAGTTAGAT	169602
Qy	1217	-----	1217

Db	169603	CCTCATAGTAGGATACGTGCTGACTAATTTCTCAGAAAGCAGGGTGACTGTTGTTATTAAAC	169662
Qy	1217	-----	1217
Db	169663	AGCAGTTAGCAACCTTGGGGTGATCCTACCTTTTATCCCCCANGGGTGCTCTCAGAAGCTG	169722
Qy	1217	-----	1217
Db	169723	TACTAATGGCCACACTCCGGTTTCACAGAGGTCACCTAAGTGTATACAATCCACAAAACC	169782
Qy	1217	-----	1217
Db	169783	GCATSCGTTGGTTTGAGAACTCCTGGGGTGGGAGCTTACTCCCTCTAAATCAGTGTGTT	169842
Qy	1217	-----	1217
Db	169843	CTTAGTTGGCCGACCTGCCCCAGGGACATCAGGGGGCATTTAAGGGTGTTGTTGGTGGT	169902
Qy	1218	-----	1222
Db	169903	TATAGTCGACAGAGGTGGAAGGGCGTGAGTCACCTAGTGATTCCAGACCAGGAAGAGTG	169962
Qy	1222	-----	1222
Db	169963	CCTGGCTGCTCTCTACCTCAGTTTCCCTGTGGTTTGGTGGCTCGGGACAGCAGTCCCAGC	170022
Qy	1223	-----	1225
Db	170023	TCTTGAAGGCTCAGGCAGGAAGGAGGCTGGTGAGTTCCAGGCTAACCTGAGCTATATTG	170082
Qy	1225	-----	1225
Db	170083	CCAGAGACCTTGTCTCTATGGAACAGTATTTCCCGATAGCGTCTCTGGGAGGTACAGTG	170142
Qy	1225	-----	1225
Db	170143	AGAGACAGGCTGTGGCAGCAGTCTGACACTGGGAAGCCAGGATGAGTGACGGTGGCAA	170202
Qy	1226	-----	1231
Db	170203	TGGTGGGGGGAAGTTACACATGATCATGCTGTGTATACCTTGTGGGTGGGAGTAGG	170262
Qy	1232	-----	1240
Db	170263	TGATGTGCAGGAGTCAACCACTCTACCCCTATAGGGCAGCCAGCAGACGTCTCTACA	170322
Qy	1240	YsGlyThrIleThrArgIleGlyGluAspSerProSerArgLeuAspArgGlyArg	1260
Db	170323	AGGTAACCATCAGCAGGATCGTTGGTGAGCAGCCCAAGTCGCTTGATCGGCGACGAG	170382
Qy	1260	luAspSerLeuProLysGlyHisValIleTyrcGluGlyLysGlyHisValLeuSerT	1280
Db	170383	AGGATACCTTGCCTCAAGGGCCATGTCATCTATAGGGCAAGAAAGGCCATGTCCTATCCT	170442
Qy	1280	YrGlu	1281
Db	170443	ATGAAAGTGGGTACAGAGGACAGACCATGAGGCACCTGGCAGTCAGGTGGGTGGCTCC	170502
Qy	1281	-----	1281
Db	170503	TGCTGTCCCTTTCTTTCCATTCTTAAGGATGTCCTTTCTCAAGACAGACATAGCAAGACT	170562
Qy	1281	-----	1281
Db	170563	CCTGAGAGCAATCCCCTTGTAGCTGGGCTCAGAGGGACATGTCTCAGACAGTACCTCC	170622
Qy	1281	-----	1281
Db	170623	CTGCCAGGTGAGTTGCCCTCGCAAGAGTGTGTGGAAACCCCTTCTGTGTAAGGAGTC	170682
Qy	1281	-----	1281

Db	170683	AGGAAACACAGACTGTAA	CATGTGTCCCCAGACAGCAGTTGTGGTGTCTATACACTTTGC	170742
Qy	1281	-----	-----	1281
Db	170743	TGAGAGGGGAAGTCCTTT	TGGCAGAGCAGAAAGCAGCCAGGGCCCTTTGGGTTGTGTG	170802
Qy	1281	-----	-----	1281
Db	170803	TGTTAGCCTCAACAGGTT	TAGGACTGGGCCCTAGGCCCTGCTGCCACCTTAGACATAGGG	170862
Qy	1281	-----	-----	1281
Db	170863	AAACGGGAGATGGACATAG	TACATTCAGTGGCCCTTCTATGAGGTTACCATCTGGCCC	170922
Qy	1281	-----	-----	1281
Db	170923	TGTTTAGACAAGCAACCTT	TGTTAGGGAGGAGCCATGGGATCTATAAAGGTGTCTTT	170982
Qy	1281	-----	-----	1281
Db	170983	AGAGCAAGGCAGAGCTGA	AGGGACACACAGGAGCAGCAAGGAGTGACTTCCTGTCTCTGC	171042
Qy	1281	-----	-----	1281
Db	171043	AGGCATACAGCCAGAGGG	GTGAGGGCTGAGGGCTGGCTTGGTGGCTTCCCTACAGAGGCTGGG	171102
Qy	1281	-----	-----	1281
Db	171103	AAGCTGGGGAATCTTTAG	TGCGTGCACAGGAAGGGCTGTGTGTAGAGCCATGACACCC	171162
Qy	1281	-----	-----	1281
Db	171163	TTCTCTGTGGTAGAAAAG	AGCCAGGGCTGCCTAAAGTTGAGGAACCCATACCTGGGA	171222
Qy	1281	-----	-----	1281
Db	171223	TGCGTCTCCAGTACCAAG	CACCACTCACTGCTGCCATGACTACTGCGAGGCTG	171282
Qy	1281	-----	-----	1281
Db	171283	GGGCTGCCAGCTTCCTAT	TTCCTTGAGTCTCAGAGCCCTCGTCTAGAGGTCACAAG	171342
Qy	1281	-----	-----	1281
Db	171343	TGGCCCATATGGGCTTTG	GCACCTGACATAAAGTACAGTCTTAAGGACAAAGGCTACAGC	171402
Qy	1281	-----	-----	1281
Db	171403	TGACCCCGGTGCTGGCTG	CATGTCCAGGTGACAGACCAAGAGCAAGGAGAGCGGTGAT	171462
Qy	1281	-----	-----	1281
Db	171463	CAGAGGTCTGGCCAGAGC	CCCTGAGTCACGGAGTGGGNGGCTCTAGAGTCTCAGGTTGTA	171522
Qy	1281	-----	-----	1281
Db	171523	ACTTTACGCCATATCTTG	GTGGAGACTGGGTATCCTGTCTCTTTTGGGGCCAGTGG	171582
Qy	1281	-----	-----	1281
Db	171583	GTCTGAAGCAGTTCAGAG	TAGGGCCACCGTTCAGGCCCCCAGCCAGCTGCCAGCCCTTTC	171642
Qy	1281	-----	-----	1281
Db	171643	TTCGCAATCTCCCATAGC	AGCATCCAAGTCAATTTGGGGATAGCAATGCCACTCTGTCC	171702
Qy	1281	-----	-----	1281
Db	171703	CCGACAGCTGAGGCTTCA	CATGGTCTTTGGACCCGGGTGCTCACACATCCCTTTGTATACA	171762
Qy	1281	-----	-----	1281
Db	171763	TGCAGACCTGCAGGGCTT	TGTGGGGCTGCTCTGCGGACTCAGACCGGTGTGACATCTTC	171822
Qy	1281	-----	-----	1281
Qy	1281	-----	-----	1281
Db	171823	GTTAGTGTGCATACACCA	TGGSCACAGACATACACCCCTTACACAGCAGTGG	171882
Qy	1281	-----	-----	1281
Db	171883	CAGGCATGCTGGCACTTT	CTGTCCAAGAGCAGTTCACAGTTGAGCTGTTTGGGGTCCC	171942
Qy	1281	-----	-----	1281
Db	171943	ATCTGTGGGGCAGGGCCT	CTCTTTCAGCTCTCCCTGTTTGGCTAACTCATCCCTGTGTGT	172002
Qy	1282	-----	-----GlyGlyMetSer	1285
Db	172003	GTCGTGACCTTCTGCTGC	CAACCATGTGCTTTTCTGGCCCTCAGTGTGTATGTCG	172062
Qy	1286	ValThrInCysSerIleGlu	AspGlyArgSerSerSerGlyProHisIleuThrAla	1305
Db	172063	GTGTCACTGCTCCAAGGA	GGATGGAGGAGCTCAGGCCACCCACAGAGACTGCC	172122
Qy	1306	AlaProLysArgThrTyrAs	pMetMetGluGlyArgValGlyArgAlaIleSerSerAla	1325
Db	172123	GGCCCTAAACGACCTATG	ACATGATGGAGGCCGTGTGGCAGGACCATCACTCAGCC	172182
Qy	1326	SerIleGluGlyLeu	-----	1330
Db	172183	AGCATCGAGGGTAA	-GTGTCTCACTGTAGTCTACCGGGTCAAGCACCATCCACGA	172241
Qy	1330	-----	-----	1330
Db	172242	GACCCAGGTACCCAGGTAA	GAATGGAAGGAGGGTGGCCCTTGACATGCTTGGAGGG	172301
Qy	1330	-----	-----	1330
Db	172302	AGCCAGTCAGTAGTTGGCT	CAATATTAAACAAGTAGAATACTGAGGCACAAACCCGGTGA	172361
Qy	1330	-----	-----	1330
Db	172362	AAATGAGCCCGAGGAAC	CGGGCAGCTTGTCTGGTGTAAAGTCAACAAGAGACTG	172421
Qy	1330	-----	-----	1330
Db	172422	TTGGTGTAGAGCCAGCCCT	CTCTGGACACCATACGTTGACACTATCCCTGTCTCCCCA	172481
Qy	1331	-----MetGlyArgAlaIle	ProProGluArgHisSerProHisIleuLysGluIn	1348
Db	172482	GGACTCATGGGCCGTGCCA	TTC-----CCTGAGCAGCAGAGCCCC--CACCTCAAGGAGCAA	172535
Qy	1349	HisHisIleArgGlySerIle	ThrGln-----	1357
Db	172536	CATCATCCGAGGCTCCATC	ACGCAAGGTATATCTCTTACGTTGTGTAAATTCACCT	172595
Qy	1357	-----	-----	1357
Db	172596	CCAGCCAAAACACAGCTAA	AGGGGTGGAAAACGAGGGTGAATCTCCCGTCCGTCTTAT	172655
Qy	1357	-----	-----	1357
Db	172656	AGTGGAGTCACTACTCTGT	ACTTGGGGCTCATTTCCAAAGCCCATGCTGCATGCCCTTGCT	172715
Qy	1357	-----	-----	1357
Db	172716	AGTGAATTGAGAGGGGTCC	CAGATTGAATGTGCTTTTAGTCTCATGCTTTGTGTAAGTCTG	172775
Qy	1357	-----	-----	1357
Db	172776	AGCAATGTGTTTCACTTAC	TAGGCTCCTCACTCATATGAAGAAACAGGCTTAGAAAAAGTCA	172835
Qy	1357	-----	-----	1357
Db	172836	GTTTGCATACCAAGGTCC	CCACAGGAAGGATTTGTGTGACCCAGGCTGAGCCAGGCAAG	172895

Qy	1358	-----GlyIleProArgSerTyrValGluAlaGl	1367
Db	172896	GCAGTCGAGAGCCAAATGATGTTGTCGGCAGGAGCATCCGAGGTCATATGTGGAGGCCA	172955
Qy	1367	nGluAspTyrLeuArgGluAlaLysLeuLeuLysArgGluGlyThrProProProPr	1387
Db	172956	GGAGGACTACTTACGTAGGAGGCCAAGCTCTTGAAGCGAAGGACCGCACCGCCCC	173015
Qy	1387	oProProSerArgAspLeuThrGluAlaTyrLysThrGln-----AlaLeuGlyPr	1404
Db	173016	ACCGCCACTCGGACCTGACTGAGACTCAAGACCGGGCCCTCGACCCCTCTGGGTCC	173075
Qy	1404	oLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgse	1424
Db	173076	CCTGAAGCTGAAGCCAACTCATGAGGTGTGTAGGACTGTGAAGGAGCGGGCCGCTC	173135
Qy	1424	rIleHisGluIleProArgGluLeuArgHisThrProGluLeuProLeuAlaProAr	1444
Db	173136	TATCCATGAGATCCACGAGAGGAGCTGCGCGCACGCCGAGCTGCCCTTGGCACCACG	173195
Qy	1444	qProLeuLysGluGlySerIleThr-----	1452
Db	173196	GCCTCTGAAGAGAGGTTCCATCAC-CCAGGTACGAGGACACAAAGAGGGCGGGGCT	173254
Qy	1452	-----	1452
Db	173255	TGGCAAGCCTGAGGGAGTTCATATGCTGCCCTCATTTGCTGTTTAAACTTAAGGCC	173314
Qy	1452	-----	1452
Db	173315	CAGAGAGTCAGCGGCTTGCTCTGGTCACTGACAGGGCTTTGTGAACCAAAGTCTTTGG	173374
Qy	1452	-----	1452
Db	173375	TCACTTGGCTGCTCCACATTTGCTACTGTCATGTTGTGTGTCAGACTAGAGCCACTGT	173434
Qy	1452	-----	1452
Db	173435	CCCCTACTTTGGTTCATGGCTTGATGGAAGAAAGAAAGACAAACAGCCAGACCACTA	173494
Qy	1452	-----	1452
Db	173495	GGGAAGCTGTGTCTACCTGAGCCAGGGACACACAGTGTGAGCAGACCAATGATTTAGG	173554
Qy	1452	-----	1452
Db	173555	GTATCTTTTGAAGAACTTAGTGTCTGCTCCTCATTTGGGCCAGCACTTGGAGGGGGAG	173614
Qy	1452	-----	1452
Db	173615	TCAGGAAGATTGTAAATTTGAGGCTAGCCTAGGCTACAAAGCTGTACTCTGGCTTACTGC	173674
Qy	1452	-----	1452
Db	173675	CGATCACCAGTTTACAAATTTAATGACACACAAACATTTGAGTGAATTTGAGATCTGG	173734
Qy	1452	-----	1452
Db	173735	GACCTAGCTGTCTACTGTCCCGAGCCTGGATCCCTGTGGGATCCTCTCAAGATGTCC	173794
Qy	1452	-----	1452
Db	173795	AATAATTTCCCGTCTCCGTGGTGTCTCTGAGAGCCACAGGAATGTGTAATCCAGCCT	173854
Qy	1452	-----	1452
Db	173855	GGGACCGGAGCCAGAGCTGGGGTGCACAGAGATGCATCTGATATGAGTCCATGAGTCT	173914
Qy	1452	-----	1452
Db	173915	GCAGGATGTCGGGGGCCACACTGGGGTGAGGAAGGTGCAATTTGACCTGTAGTTAGCCACGC	173974
Qy	1452	-----	1452
Db	173975	TCCTGACTCCCTCCACAGTGGGGTCTCTGGAGCTATAGCCAGCTGTAAGAGGCTGAGAG	174034
Qy	1452	-----	1452
Db	174035	TAICTAGCCTCTGACGTCTGTCTTGTCTCAAGGGGTTAGAGATGCCCATCTCTGTGCAGAC	174094
Qy	1453	-----GlnGlyTh	1455
Db	174095	ATACAGGAGAACAGCATCTGGCCAGCCACATACAGACCTGGACCCCTCTCCCCAGGGAC	174154
Qy	1455	rProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgse	1475
Db	174155	CCCACTCAAGTAGCACTCCCGGGTGCCTCCAGTGGCTTCCAAGAAACACGACGTACGCTC	174214
Qy	1475	rLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAs	1495
Db	174215	CATCATCGGAGCCCGCGCGGCTTTTCTTGCCTGCACCCACTGGACATAATATGGCTGA	174274
Qy	1495	pAlaArgAlaLeuGluArgAlaCysTyrGluGlySerLeuLysSerArgProGlyThrAl	1515
Db	174275	TGCCCGGCACATAGAGCGCGCTGCTATGAAGAGAGTCTGAAGAGCCGGTCAGGAGCACAG	174334
Qy	1515	aSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLys	1535
Db	174335	CAGTGTGTGAGGGGCTCCATCACACGCGGGGCCCCAGTTGTCTGCTGCTGAGCTAGGCAA	174394
Qy	1535	sProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuPr	1555
Db	174395	GCCTCGGCAAGGCCACTGGCTTACGAGGACCAACGCGGGCACCTTACACGCCACCTGCC	174454
Qy	1555	oArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSe	1575
Db	174455	CCGAGGCTCACTGTGACTTACAGGAGGCCACGCCACGCTTCAGGAAGGTGAGGC-TC	174513
Qy	1575	r-----	1575
Db	174514	CCCTGGGCTGCCAGCTGTGGGAGGGGACAGCTCATTTTCCACGCGCAGTCTGTCTGCTG	174573
Qy	1575	-----	1575
Db	174574	CCCTGCTCCACACTGTCATACAGAGCCCCCGTGTCTTCCCTTACCAAAATATTTTAA	174633
Qy	1575	-----	1575
Db	174634	TAATTCAAATCAGGGCTGAGGAGGGGCTCAGTAGCTTAAAGTCTTGTGTACCAGCATGA	174693
Qy	1575	-----	1575
Db	174694	GACTCAGAGTTCAAAATCCCGGTATCCACATAAGCCAGGCACTCATACACATCTGTAA	174753
Qy	1575	-----	1575
Db	174754	TTCTCATGTTCACGTGAGACAAAGGCTAAGACAGAAATCTCTCTAAAGGTTATAGGCCA	174813
Qy	1575	-----	1575
Db	174814	TGAGTCGGGTCTCAAACATGGCGGAAAGTGGATGACACCTCTCCCCCCCCCTCTATT	174873
Qy	1575	-----	1575
Db	174874	TTATACACAGCATAGCACACTCATATCAGACACTCGCGCTACATTTCTGTGTGCAC	174933
Qy	1575	-----	1575
Db	174934	ACATGTACACAAAAACAATAGTTTAAATGAAGCAGTTCCTTCTGTAGGCTTACCT	174993
Qy	1575	-----	1575
Db	174994	CTCCAGTAGCGCTGCTGGTCTGAAGCACCTGCTGTGACGTGACCGCGGCATCTCTCTCT	175053
Qy	1575	-----	1575

Db	175054	GAATCCGCTTCTGATTCTCTCGCATTTCTGACCTCCGCTCTTTCTCCGATCCTCCACAT	175113
Qy	1575	-----	1575
Db	175114	GTGAGCCTTTCCGTGTGCTGGCCCTCTATGCCGAAGGCTCCACCTTTGGTTTGTCTCT	175173
Qy	1575	-----	1575
Db	175174	AGCCAACCACTGTATACCCACGTCACCATGTGAGGGCTCGAGAGGCTATAGGCACC	175233
Qy	1575	-----	1575
Db	175234	TCCAACCTAGTTCCAGCCACATCCGGCCGGATTGTGTAGGGTGTGAGACCAGTGT	175293
Qy	1575	-----	1575
Db	175294	AAITTCATTTGATGAGGAGCCCTTCCAGAGCACTGAGATCCTCAACATCTCTCTCT	175353
Qy	1575	-----	1575
Db	175354	CAGCCCGTCGACTTCTACTCTGGGTTCTATCTCTTATCTGGAACAATGGCTCCTTGGAG	175413
Qy	1575	-----	1575
Db	175414	CTCAGGGTCTTGCCCTCTTCTCTCAGGGTGTCTCCTCTCTCTCTCTCTCTCTCTCTCT	175473
Qy	1575	-----	1575
Db	175474	AGCCCATGATTTAAGTCAGACTGGAATGAATAGGGCAATGGTACTGGGGAGTAATGAA	175533
Qy	1575	-----	1575
Db	175534	TGCTTTCAAGGAGCCAGGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	175593
Qy	1576	-----	1576
Db	175594	CCTGGTGGCCCGTTGAGGCTGAGGCCACTTTGACCTGTCCCCACAGGAGCCTCCTCTC	175653
Qy	1576	-----	1576
Db	175654	rSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerPr	1596
Qy	1576	-----	1596
Db	175655	CAGCAAGCGCTCCAGAGCCGGAAGCTGACCTGCACACCCCGGAGATTCCCAAGTCCC	175713
Qy	1596	-----	1616
Db	175714	oHisSerThrValProGluHisProHisProHisProHisProHisProHisProHisPro	1616
Qy	1596	-----	1616
Db	175715	gGlyValSerGlyValAspLeuTyArgSerHisIleProLeuAlaPheAspProThrSe	1636
Qy	1596	-----	1636
Db	175716	GGGTGCTGCTGGTGTGACCTGTACCGTGGCCACATCCCATTTGGCCCTTGACCCACCTC	175833
Qy	1636	-----	1644
Db	175834	CATACCCGAGGATCCCTCTGGAAGCAGGTGGGTGCTGGGCTGTACATGTGGGGCTT	175893
Qy	1644	-----	1644
Db	175894	ATTGATGTGCTTCTTGGGTGGACACTTGGGTATGGCTTAGTGTGCCATCACCAGCT	175953
Qy	1644	-----	1644
Db	175954	GCTTACCCAGGAGGCTTCTAGGATTTGGCTTTTGTGGTGTCTCTCACCCTAGGACT	176013
Qy	1644	-----	1644
Db	176014	CATCTGCTCCTTGCAAGTCAGTAATAGAGGGCTGCAGTGCAGTGCATGATGGAACC	176073
Qy	1644	-----	1644
Db	176074	TTTTTAAAGGCTCCTTTGTGAGGTGTAGCACCTGTATTTTCGATGCTATATTATGGGG	176133
Qy	1644	-----	1644
Db	176134	GCACTTCCGAAACATTTAAAGCAGGCCCCCACATGATGAGGAGGAGTCTTTAAGCTC	176193
Qy	1644	-----	1644
Db	176194	TGTGTGTGCTTGGGAGGCATCTGGTGTACAAGGAGAGAGCTGTTGTGCGAGAGT	176253
Qy	1645	-----	1645
Db	176254	GAGGAAGTGAAGATAAGTCTATGGGCCACTGACCCCTCCATCCTCCCTCCCTCCACG	176313
Qy	1645	-----	1645
Db	176314	aAlaAlaAlaTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy	1665
Qy	1665	-----	1665
Db	176315	AGCCGAGGCTACTACTCTCCCGGCACTTGGCCCGGCACTTGGCCCGGCACTTGGCCCGG	176373
Qy	1665	-----	1665
Db	176374	rProProTyLeuIleArgGlyTyProAspThrAlaAlaLeuGluAsnArgGlnThrIl	1685
Qy	1665	-----	1685
Db	176375	CCGCTTACCTCATCCGGGCTACCTGCACCCGGGCGCTGGAGAACCGCCAGACCAT	176433
Qy	1685	-----	1705
Db	176434	eIleAsnAspTyTyIleThrSerGlnGlnMetHisHisAsnThrAlaAlaAlaMetAlaGl	1705
Qy	1705	-----	1705
Db	176435	CATCAATGACTACATCACCTCGCAGCAGATGCACCAACGCGAGCTCTGCCATGGCCCA	176493
Qy	1705	-----	1725
Db	176494	nArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerSerLeuAlaLeuAsnTyAl	1725
Qy	1725	-----	1725
Db	176495	GCAGCGGATATGCTGAGGGCTGTGTACCGAGAGAGTCTCGCTGGCCCTCAATTATGC	176553
Qy	1725	-----	1729
Db	176554	aAlaGlyProArg-----	1729
Qy	1729	-----	1729
Db	176555	CGCTGGCCCGCTGGTGTGAGTTGCTGGGGCTGTGCCATCAGCAGAGACTTGTCTCTGGC	176613
Qy	1729	-----	1729
Db	176614	TGTGGTGTGCAAAAGAAGAGTGCACACGTGTGCACTGTGCAGCTGGCGGAGGGGCACAGTT	176673
Qy	1729	-----	1729
Db	176674	GGCTGGGAGGACTAGTGTGTGCTGCTGCTGGGGAGAGGAGAGAGAGAGAGAGAGAGAG	176733
Qy	1729	-----	1729
Db	176734	GTATCTGTTCACTCTAGAACTGTGTGTAGCAGACACTCTTGCCACCATGCACCGTGGGT	176793
Qy	1729	-----	1729
Db	176794	CTTTGCGTGTGTCTGAAACCTTGTATGGGCATCTTCTGTCTATCCAGAGATCAAAA	176853
Qy	1729	-----	1729
Db	176854	TACAGGCTGATGTGGCACTTGCACGCTATATCTTAAGTACGAGAGAGTGCAGAGCCAG	176913
Qy	1729	-----	1729
Db	176914	CCTCAGCTACATAAATTTTGGAGCCAGCCTGGGCTATATAGTCTGCCTCTATACAAATG	176973
Qy	1729	-----	1729
Db	176974	ACTGAAAAGAAAGAAATGAGAGCATGCGGCTTGTCTCAGAACACTGACTTCTGTGAAACAGC	177033
Qy	1729	-----	1729
Db	177034	CTGGGCTCACTTTGCTCGACTGTTTCAGAGATAACCTTGTGGCCCCAGTTAGTATACGGA	177093
Qy	1729	-----	1729
Db	177094	AGCCACTCTGGGGTTTTAGAGATGGGAGCCCAATGCAAGGAGCTGATCTGTGGGAAG	177153
Qy	1729	-----	1729
Db	177154	GCAGTGTGTTGGGTCAAAAGATGTTGTTTGAGAGACACTGCAGTCTATAGCCTTGAGAA	177213
Qy	1729	-----	1729
Db	177214	GGGGCTTGGGGCCACTCTCTGGCACCCAGGAGGGGTAGTGAGGTGAAGGTCTCAGGCCT	177273
Qy	1729	-----	1729

Qy	1730	-----G	1730
Db	177274	GCACCTTTGGGACAGGACATCAGGCACCTACAGCTGACATGCCACCCACCCCTCGAC	177333
Qy	1730	lyllelleAapLeuSerGlnValProHisLeuProValleuValProThrProGlyT	1750
Db	177334	GCATTATCGACCTGTCTCCAGTAGTGCACACCTGCGCGTGTGTGCGCCCACTCCAGGCA	177393
Qy	1750	hrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerS	1770
Db	177394	CCCTGCGCACCGCATCGACCGCTTGCCTACCTCCCACTGCGCCCACTCCATTCAGCA	177453
Qy	1770	erArgHisSerSerProLeuSer-----	1778
Db	177454	GGCTCACAGCAGTTTCACTGCTCCCAAGTGGCGCTGCCAATGCTGGTGGGTCT	177513
Qy	1778	-----	1778
Db	177514	GTGCTCACCCCTGTGCCCTGCAAGTAAGTTCTGGGCCCACTGGGCTCTGGCTCACACTC	177573
Qy	1779	-----ProGlyGlyProThrHisLeuThrLysProThrHisSerSerS	1794
Db	177574	TGGCTTGGGGTTTCCAGGAGGCCCACTCACCTAGCTTAACCAACTGCCACGTCATCAT	177633
Qy	1794	erGluArgGluArgAspArgGluArgAspArgAspArgGluArgGluLysSerI	1814
Db	177634	CGGAGCGGGAACGGGAACGAGAGCGAGAACGAGAC-----AAGTCCA	177675
Qy	1814	leLeuThrSerThrThrThrValGluHisAlaProLleTtpArgProGly-----	1830
Db	177676	TCCTCACGTCACCACTACAGTGGAGCATGCACCCATCTGGAGACCTGGTAGGSCACCCC	177735
Qy	1831	-----ThrG	1832
Db	177736	AGACCCCGCCCAACCCAGGTCTCATGGGCCACCTGGGGCTGCCCTGACTCGGCGG	177795
Qy	1832	luGlnSerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProA	1852
Db	177796	CCTG-TCAACGAGTACGAGCAGAGCAGCGGGCTGGGGG-----AGCAGCGCGCCCG	177848
Qy	1852	laSerHisSerHisAlaHisGlnHisSerProLleSerProArgThrGlnAspAlaLeuG	1872
Db	177849	CTTCCCAACAC-----CACACGACTCGCCCATCTCCCGCCGACCCAGAGCGCTTGC	177902
Qy	1872	lnGlnArgProSerValleuHisAenThrGlyMetLysGlyLlelleThrAlaValGluP	1892
Db	177903	AGCAGAGCCCGAGTGTCTGCACAAACAGAGCATGAAGGGCGTGTCACTCCGTGGAAC	177962
Qy	1892	roSerLysProThrValLeu-----	1898
Db	177963	CCGSCACGCCACCGTCTCT-GAGGTGGGCCAGGTTGGCACCCAGGGGGAACAAGCCTGGGG	178021
Qy	1898	-----	1898
Db	178022	GGAGGGTACACAGATGGATTGTCTGAGAGGAGGGGTGTGGGGGTGGGCGAGATGGGTAGG	178081
Qy	1898	-----	1898
Db	178082	GAATGGAGCAGATGGCAGGATCGGAAGGTGGGAAAGTGAGAGACTATGAGAGTGGGGA	178141
Qy	1898	-----	1898
Db	178142	TGGGTGCTTGTGAGCGGGGAGGATGGCCAGGGGAGAGGGCACACAGTGAGTGGACG	178201
Qy	1898	-----	1898
Db	178202	GGAGATAGATGACAGATAACGAGCAGATGACAGTGGGTGGATAAAGGATGGATGGATGGA	178261
Qy	1898	-----	1898
Db	178262	TGGATGGATGGATGGATGGGTGGGTGGGTGGATGGATGGATGGATGGATGGATGGGAG	178321
Qy	1898	-----	1898

Db	178322	CAGATGGACGAGTGGATGGATGGACTAATGGATGGATGGACAGACAGACAGCAGGAGG	178381
Qy	1898	-----	1898
Db	178382	CAGACAGATGAACGAGTGGATGGATGGACTGGCGGATGGATGGACCGACCGACGATGGA	178441
Qy	1898	-----	1898
Db	178442	TGGATGGACAGGACAGACAGACAGCAGACAGATGAACGAGTGGATGGATGGATGGG	178501
Qy	1898	-----	1898
Db	178502	CGGATGGATGGACCGGACCGGATGGATGGATGGACAGGACAGATGGACGAGTGGATGGA	178561
Qy	1898	-----	1898
Db	178562	TGGACTGATGGATGAATAGACAGACCGGACGCATATATGAGTAGACAGCAAGTTGATGGG	178621
Qy	1898	-----	1898
Db	178622	GGCGGATGAATGGGTGGATGGATGGATGGATGGATGGAACAGCTGGATAGACAGGT	178681
Qy	1898	-----	1898
Db	178682	AGACAGACAGATGAATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	178741
Qy	1898	-----	1898
Db	178742	NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	178801
Qy	1898	-----	1898
Db	178802	NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	178861
Qy	1898	-----	1898
Db	178862	NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	178921
Qy	1898	-----	1898
Db	178922	NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	178981
Qy	1898	-----	1898
Db	178982	NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	179041
Qy	1898	-----	1898
Db	179042	NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	179101
Qy	1898	-----	1898
Db	179102	NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	179161
Qy	1898	-----	1898
Db	179162	NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	179221
Qy	1898	-----	1898
Db	179222	NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	179281
Qy	1898	-----	1898
Db	179282	NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	179341
Qy	1898	-----	1898
Db	179342	NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	179401
Qy	1898	-----	1898

Db 183766 ACTCAACACCCATTAACCGGAATGAGCCTGAATACAGTAAGGAGAAGCAAGCCAGCAGCA 183825
QY 2304 ----- 2304
Db 183826 GGCACCTGGGAGACAGAGTGCTCTTAAGCCTTCTGTGCGAAGGGCTTGATAGTCTGCCAGT 183885
QY 2304 ----- 2304
Db 183886 GGATACCGACGGCATATAGCACCTCAGTAATGCCCTACGCACAGTAGTGCTCAGCGAAA 183945
QY 2304 ----- 2304
Db 183946 CCTATAGAGACAGCCCTGTGCTGGCGCGAGAGACAGTGCCTACTCCAGGTTTCATGC 184005
QY 2304 ----- 2304
Db 184006 TGCAGGCATGGCGCTTTCTCCACAGCTGATGTGGGTGGAGCAAGACCCTTGTCTTAGC 184065
QY 2304 ----- 2304
Db 184066 TCAGGTGCGAGGACCATATGCTTATGAGATGCCACTGGGAAACACACTTGGAGGCTT 184125
QY 2304 ----- 2304
Db 184126 CGCACTGTCCCTCCACAGCAGCCCTCCCTAGGTGGCTTGGACAAATCCTTCTGG 184185
QY 2304 ----- 2304
Db 184186 AGTTAGACTACAGAAGCCCTTGCGAGTTAGCACAGGCTTAGGTAGTGACTAGCCAG 184245
QY 2304 ----- 2304
Db 184246 AGTAAACCTGGAGCATCAGATACCTAGTTGTGCCAGACATTC AACCCAGACCGGTACAGGG 184305
QY 2304 ----- 2304
Db 184306 AGCTAACGGGTGCAGATCTCCAGAACCTGTGAGCTCAGGCTTACCCACTTGTACCC 184365
QY 2305 ----- AsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyTh 2322
Db 184366 TGCTCTAGATATGGCCAGCCTGGACGGAAATCTTCAACATGCCGCCATCACTGGAGC 184425
QY 2322 r----- 2322
Db 184426 -AGTTAACCTCCGCTGTGTCGCGATAGCAATCTCCCTTCCCACTACTGACTTAGTGA 184484
QY 2322 ----- 2322
Db 184485 CAAGGAGCCTTAGCCAGTTCAAAATTTGAGCCCTCTCCCTGGCCTTAGTGGACCCAGAAATG 184544
QY 2322 ----- 2322
Db 184545 GCACAGAGCCAGCCTCCACTGCCTCGAGTCTCCCTATACATGAGGAGCGTCCAAGGTTT 184604
QY 2322 ----- 2322
Db 184605 GAGATGAGGTTTCTCCGGTGGGTGTAGGTACACCTCGTAATACTTCCCAATTGCT 184664
QY 2322 ----- 2322
Db 184665 GGGGCTGAGGACGTCACCCAGTGTGGGCCCTAGGAAATTTGGCCTTGGGGTCCAGAGCTTC 184724
QY 2322 ----- 2322
Db 184725 AGGTGCCACAGGCATCTTCACACTACATCTCCATTTGGGGTTCAGCAGAGAGCCCT 184784
QY 2322 ----- 2322
Db 184785 GGCCTTTTCTAAATCAGATGCAAAACGGGTAATGTGAGAGAAACATGTCACACTCAGGTG 184844
QY 2322 ----- 2322
Db 184845 CAGTTTGTGCAGAAAATTTTCAGGCATTTACAGTGTGGTGTGGGACTCAGAACATCCGGCTT 184904

QY 2322 ----- 2322
Db 184905 GCCATGGTTATGATGAGCTTGGCCCTGGGAAAGAAAGCTGTCATGGGAATAAATGGCTCAG 184964
QY 2322 ----- 2322
Db 184965 GAACACCTGAGCCGGGACTGTGACACTGGGGCAGCTGTGAGGACTTGCACCTCCAGGCTGG 185024
QY 2322 ----- 2322
Db 185025 AGGAACAGGCAAGGGTTCACTACCTTCTTATGTGTGAGCTTTGGGCCCCAGGGAATAAG 185084
QY 2322 ----- 2322
Db 185085 GTGGCCAGATACACCCCTTGGGAACATTTCCAGAGGACCACAGACCTGCAGAGGGGGTGG 185144
QY 2322 ----- 2322
Db 185145 GGGTGGGATCCTGGCAGGGGCTGTACAAAACCCAGAAACATACTGGGGACTTTTCTGAGC 185204
QY 2322 ----- 2322
Db 185205 CTCGGGCAGAGGAAACTTTTGACAGTTGAAGATGAATGTTCTAGAATGTTCTAAATCT 185264
QY 2322 ----- 2322
Db 185265 TGATGAGTTCTAGACACCTGGGTGTGTGCAGGTAGAGATTCACAGGGTAATACACGC 185324
QY 2322 ----- 2322
Db 185325 ACACACACACACACGCTGCTCTCAAAACCCCAATAAATGTTAGTCTCTCAAAGT 185384
QY 2322 ----- 2322
Db 185385 CATGTTTCCAGGCATGCCAATGACACTCAACATCGGGCCAGAGTAAATGAGGAAGGTC 185444
QY 2322 ----- 2322
Db 185445 ATGCCTGTGCCAGTACGCGCCCTCAGGTCACAAGGCCAGGAGCGGTCAGTCTCCATC 185504
QY 2322 ----- 2322
Db 185505 AGGAAACTCCCTCCACCCCTCTTCCCTGTACCTCGTATCTCCCTCTGTATTTCTG 185564
QY 2322 ----- 2322
Db 185565 GATGTGAGGAGTGGGTGGGGCTTGTATTATTTAGGTTAGGGCGTGGAGGGTGTGTGTG 185624
QY 2322 ----- 2322
Db 185625 TGTGCGTGGTGGTGGCTGTGGCATGCAATTTACCCCTGTCAAGAGTTCCGGCTTTT 185684
QY 2322 ----- 2322
Db 185685 TCTTGTGTAAATTTGCTTAGCACTAGAGGAAACGTGAGCATCCCTGACCGAGCTTGCCCG 185744
QY 2322 ----- 2322
Db 185745 AAACCCAAACTGAGGCTCCACAGCCACCCAGAACAGGATTTAGGAAGTGCATGTCTCCTTC 185804
QY 2322 ----- 2322
Db 185805 CATCCTCAGAGCCTGGCTATTTTACATTTGCCCGCATTTTCAGCCATGTGATTTTCACAG 185864
QY 2322 ----- 2322
Db 185865 GTTTGTGAGCTGAGGTTTTCACGGGGCACAGAAACGTGCCAGTGGCATAGGTGGGAAT 185924
QY 2322 ----- 2322
Db 185925 CCTGCGCATGACGTCTGTCTGTGTTTTTTTGTGGTGGTGATTTGGGGTTTTTTATTG 185984

2243 ProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerPro 2262
|||||
Db CCACTGCTGTATCAGACGGGGAACAGCGGAGGCC---AGGATGGGCTCTAAGTCTCCA 1125

2263 GlyAsnThrSerGlnProAlaPhePheSerLysLysLeuThrGluSerAsnSerAlaMet 2282
|||||
Db GGCAACACCGACCGGAGCGGCGCTTCTTCAAGTAACTGATGAGAGCAATCGCGCCATG 1185

2283 ValLysSerLysGlnGluLeuAsnLysLysLeuThrHisAsnArgAsnGluPro 2302
|||||
Db GTGAAGTCGAAGACGAGAGATCAACAGAAATCTCAACATGCCCGCCATCGAGCCA 1245

2303 GluTyrAsnLysSerGlnProGlyThrGluLeuPheAsnMetProAlaIleThrGlyThr 2322
|||||
Db GAATACAAATATGGCCAGCGCTGGACGGAATCTTCAACATGCCCGCCATCGAGCCA 1305

2323 GlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeu 2342
|||||
Db GGCCCTTATGACCTGTAGAAGCGGCGGTGCAAGAACACGCGCAGCACCAATCGGGCTA 1365

2343 GluAlaIleLeuArgLysAlaLeuMetGlyLysTyrAspGlnTyrGluGluSerProPro 2362
|||||
Db GAGGCCAATATTAGAAAGGCACTCATGGTAAATATATCATGCTGGGAGAGCCCGCCCG 1425

2363 LeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuPro---AlaAlaMet 2381
|||||
Db CTCGGCGCCCAATGCTTTTAACTCTGATGCGCAGCGCCAGTCTGCCCGTGTGCTATG 1485

2382 ProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGly 2401
|||||
Db CCCATAACCACTGCTGACGACGCGAGTGACCCACCACTCACCTCGCCAGGTGGAGTGGG 1545

2402 LysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeu 2421
|||||
Db AAAGCCAAAGGTCTCTGGCAGACCTAGCAGCGGAAAGCCAAAGTCCGACGACCAAGGCTA 1605

2422 AlaSerGlyAspArgProProSerValSerValHisSerGluGlyAspCysAsnArg 2441
|||||
Db GCGTCCGGACCGACCCCTTCTGCTCTCTCAGTACTACTCAGAGGGGAGCTGCATCGC 1665

2442 ArgThrProLeuThrAsnArgValTyrGluAspArgProSerSerAlaGlySerThrPro 2461
|||||
Db CGAACACCACTCAACCAACCGTGTGGGAGACCGGCCCTCATCTGCGAGGTCCACGCCA 1725

2462 PheProTyrAsnProLeuIleMetArgLeuGlnIleGlyValMetAlaSerProProPro 2481
|||||
Db TCCCCCTACAAACCTTTGATTTATGAGGCTACAGCAGGTGTCTGGCTCCCGCCGCCA 1785

2482 ProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaTyrAspGluGlu 2501
|||||
Db CTTGGCTTGGGAGCAGCAGCGGCGCCCTAGTGTCTCCACCAACCGCTGGATGAGGAG 1845

2502 ProLysProLeuLeuCysSerGlnTyrGluThrLeuSerAspSerGlu 2517
|||||
Db CCCAAGCCACTGCTGTGTCTACAGTATGAGACACTCTCGGACAGCGAG 1893

RESULT 20
AC073916/c
LOCUS
DEFINITION Homo sapiens 12 BAC RP11-408I18 (Roswell Park Cancer Institute
Human BAC Library) complete sequence.
ACCESSION AC073916
VERSION AC073916.41 GI:29293998
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 205283)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratingue,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,

Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.P., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S.,
Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P.,
Francz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N.,
Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P.,
Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M.,
Holloway,C., Hollins,B., Homsai,F., Howard,S., Huber,J., Hulyk,S.,
Hume,J., Ioshikes,I., Jackson,L.E., Jacobson,B., Jia,Y.,
Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U.,
King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N.,
Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O.,
Lieu,C., Liu,J., Liu,W., Loulseghe,H., Lozado,R.J., Lu,X.,
Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P.,
Marandel,I., Martin,R., Martindale,A., Martinez,E., Massey,E.,
Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S.,
Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T.,
Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M.,
Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokwenkwo,S., Ogih,M., Okunolu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Fu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shim,C., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Umani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Woodson,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kuchelapati,R.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 205283)
Worley,K.C.
Direct Submission
Submitted (05-JUL-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 205283)
Worley,K.C.
Direct Submission
Submitted (31-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 205283)
Worley,K.C.
Direct Submission
Submitted (01-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 205283)
Worley,K.C.
Direct Submission
Submitted (06-MAR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
6 (bases 1 to 205283)
Worley,K.C.
Direct Submission
Submitted (15-MAR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
7 (bases 1 to 205283)
Worley,K.C.

[illegible]

QY	1131	lGln	-----	-----	-----	1132
DB	64310	ACAGGTTTCTGGGTGTGACACAGTGTATACATACAAAGCCGTGTACTGCTGCACACCTGGT	64251			
QY	1133	-----	-----	-----	-----	1137
DB	64250	ACATGTGTGTACATGACACCCACTTAGTGTCTCAGCAGCCAGGCTCAGTGCATGTGCCAC	64191			
QY	1137	rSerGluHisAlaAlaProValGlyProValThrMetGlyLeuPro	-----	-----	-----	1153
DB	64190	ATCCACTCTGCACACACAGACCCCGTCAGTGCATGCTGCTCTCTGCTGCCACAC	64131			
QY	1154	-----	-----	-----	-----	1158
DB	64130	GTCCACACACTGCTGATATCGGTGCACACGGTGTCTCTGCTGCCCGCACCTCCACT	64071			
QY	1159	-----	-----	-----	-----	1163
DB	64070	CTGCTGTATCGGTGCACCGGTGTCTGCTGTGCCACATCCACACTCTGCTGCTGATGT	64011			
QY	1164	-----	-----	-----	-----	1176
DB	64010	ATCGGTGCACGTGTCTCTGCTGCCACATCCACACACTGCTGATCGCTCTTGCTC	63951			
QY	1176	yGlnAla	-----	-----	-----	1189
DB	63950	ACGTGCATGTGTGCACCTTGTTTCACGCCCATGTTCGGCACCCGTGGGTGTGCACACAGA	63891			
QY	1189	aGlnGluAlaSerValLeuArgGlyThrAlaLeu	-----	-----	-----	1207
DB	63890	CTCACGTGCTCATGTGTGTACAGGCACACCTTGCTTGCACAGACAGACACCTTGGAGGC	63831			
QY	1207	rIleThrLysGlyLeProSerThrArgValProSerAspSerAlaIleThrTyArgG1	1227			
DB	63830	TAGGGTACAGGGTGCAGGCAGCTGCCGTGCATCTCCACCGTGCACACCCCTTGGAAAG	63771			
QY	1227	ySerIleThrHisGlyThrProAlaAspValLeuTyLysGlyThrIleThrArgIle1	1247			
DB	63770	CTCCTGG	-----	-----	-----	63744
QY	1247	eGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSer	-----	-----	-----	1262
DB	63743	GGGCTTCAGGCTGCCAGCTCCCTCTGAGAGCCTCGCTGTGAGGCTTTGTGACGGGC	63684			
QY	1263	-----	-----	-----	-----	1270
DB	63683	AGCATGAAGCACTGTGGTCTCTGCTACACCGCTCTGCTGTCTGCTGCCACCCG	63624			
QY	1270	rGluGlyLysGlyHisValLeuSer	-----	-----	-----	1285
DB	63623	GTCTGGGCCGTGCCAGCTCTCTACCCAGTCTTGTGTGGTTTCAGGTGGCATGTC	63566			
QY	1285	rValThrGlnCysSerLysGluAspGlyArgSerSerGlyProProHisGluThrAl	1305			
DB	63563	TGTGACCCAGTGTCTCAAGGAGGAGCGCAGAGCAGCTCAGACCCCCCATAGACGCG	63504			
QY	1305	aAlaProLysArgThrTyraAspMetMetGluGlyArgValGlyArgAlaIleSerSerAl	1325			
DB	63503	CGCCCCAAGCGCACCTATGACATGATGGAGGCCGCGTGGGCAGAGCCATCTCTCTCAGC	63444			
QY	1325	aSerIleGlu	-----	-----	-----	1328
DB	63443	CAGCATCGAAGGTGATAGCAGGAGGAGACTTCATCTCTGGTGCCTCTGTTGGCGGT	63384			
QY	1328	-----	-----	-----	-----	1328
DB	63383	GGGGGATGCTGACCCCGTTTTTACAGATGGGGAACCGAGGCTGGGCTTTCTGAGGCTC	63324			
QY	1328	-----	-----	-----	-----	1328
DB	63323	CATCTGGAGGTAGCGGAGGAGACCTTCCCGTGTGGTCTCTACTCCACCATCATCTGGGG	63264			

QY 1328 ----- 1328
Db 63263 ATGACCACTGGCGGTGTAAACACTGACCCCTGTCTACTCCAGTGTGCGCTCAGGGAGC 63204
QY 1329 ----- GlyLeuMetGlyArgAlaIleProProG 1338
Db 63203 CACGGAAATGATGGCCTCACCTCTCTACCCAGGTCTCATGGCGGTGCATCCCGCGG 63144
QY 1338 luArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlySerIleThr----- 1356
Db 63143 AGCGACACAGCCCCCACCACCTCAAGAGCAGCACACATCCGCGGTCCATCACACAAG 63084
QY 1356 ----- 1356
Db 63083 GTACTGCCCTGTTCCTGTCTCCTGTGCTGCCCAACGGGTGTACAGTCACGAGGGCGC 63024
QY 1356 ----- 1356
Db 63023 GCGAGGGAGAGACACAGCCAGAGTGTGTGGGAACTCAGGACAAGTATGCAAGAGGCT 62964
QY 1356 ----- 1356
Db 62963 GCAGCACATGTATCATGTATCATGTACGTACGTACGAGCAAGGTGTCTCTTACTTATT 62904
QY 1356 ----- 1356
Db 62903 TCAAAACAAAAACAAAGCAAAATACCACCGATCACCCCTGTGCTTTAGGTGTACAT 62844
QY 1356 ----- 1356
Db 62843 AGCAACTGTCTGTGCTTGGCACTAACCCAGGTGCCACCTGCTATCGTTTACAGAA 62784
QY 1356 ----- 1356
Db 62783 TCCTGTGAGGCACATCGATTTGGGAGAGGCTTGGAGACTCCGAGAACTCTTTCAGGTT 62724
QY 1356 ----- 1356
Db 62723 CTCGCGGTGTCATGGCACAGCCAGCTACTGTGAACCTTGGCAGCTTTGTGGGTTTAT 62664
QY 1356 ----- 1356
Db 62663 TTTTATTTTTTATTTTGTGTTGTTGCTGTTGTGAGACAGGGTCTCATCTGTGC 62604
QY 1356 ----- 1356
Db 62603 CCAGGCTGGAGTGAGTGGCAGCATCTCAGCTCACTGAGCCTCCGCTTTCAGGCTTAA 62544
QY 1356 ----- 1356
Db 62543 ACAATCTCCCACTTCAGCCTCCCAAGTAGCTGGGACCACAGATGCACACCACATCT 62484
QY 1356 ----- 1356
Db 62483 AGCTAAATTTTGTATTTTGTAGAGTGGGTTTACCATGTTGCCAGGCTGTCTCGA 62424
QY 1356 ----- 1356
Db 62423 ACTCTGAGCTCAAGCTGTGCTGCGCGAGCCCCCAGAGTGTGGATTACAGCGGT 62364
QY 1356 ----- 1356
Db 62363 GAGTACTGCACCAGCCTGTGGTTTGTAGCTTCATGATTTTCATAGTGTCCCGATTGCT 62304
QY 1356 ----- 1356
Db 62303 GAGGTGTTTCAATTAATTTCTGTGTTTATGTGTGAAGACTGAGGCCAGAGGTCA 62244
QY 1356 ----- 1356
Db 62243 GATTTCTGGTCAAGGTCAACAGCAAGTGGGATTGAACTCAGGCAGACTAGTCCAG 62184
QY 1356 ----- 1356

Db 62183 AACCCACTGGTGTGAGGCTCTTGATGGTCTGGTGGCGGGCGGTGAGGTCAGTGC 62124
QY 1357 ----- GlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgA 1373
Db 62123 TGTCCGCCCGCAGGGATC-CCTCGGTCTACGTGGAGGCACAGGAGACTACCTGCGTC 62065
QY 1373 rgGluAlaLysLeuLysArgGluGlyThrProProProProProSerArgAspL 1393
Db 62064 GGGAGCCAGCTCTAAAGCGGGAGGCGACGCTCCGCCCCCACCCTCACGGGACC 62005
QY 1393 euThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluG 1413
Db 62004 TGACCCAGGCTTACAAGACGAGGCTTGGGCCCCCTGAAGCTGAAGCGGCGCCATGAG 61945
QY 1413 lylLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluL 1433
Db 61944 GCCTGTGGCCACGTGAAGAGGCGGCGCTCCATCATGATCCCGCGCAGGAGC 61885
QY 1433 euArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThr- 1452
Db 61884 TCGGGCACACGCCCGAGCTGCCCTGGCCCCGCGCTCAAGGAGGCTCCATCACGC 61825
QY 1452 ----- 1452
Db 61824 AGGTATGGCCAGGCGCAGGCACACGGGCCAGTTCTAGGAGGGTGGCGGTGTGG 61765
QY 1452 ----- 1452
Db 61764 GGCATGCCCCCTGGGCTCTCCACATGGGGAAACCGAGGCTGAGAGCCCTCGCGTACTTA 61705
QY 1452 ----- 1452
Db 61704 CAGTACCAGCTGCTCATCACGGGCTCAGCTGTGCTGTTCAGGCTGCGAGGGG 61645
QY 1452 ----- 1452
Db 61644 GCACAGGCTCCTGACCTGATTCTACTGAACCTCACATTTGTTCCCATTTCTTCAGGGAGGA 61585
QY 1452 ----- 1452
Db 61584 AACTGATCCAGAGAGGCCAGGAGCTTCCAGGCCACAGACTAAACATAGTACGA 61525
QY 1452 ----- 1452
Db 61524 GTAACTGCTCCGTGTAATCTTTGTAGGGTCCAGGTGCGGCTGAGGATGTTGCATGC 61465
QY 1452 ----- 1452
Db 61464 TTCAATTTTTCACCCCTAGCAATGCTCTGAGGTGCTTTTCTTAATGACCTTATTTATT 61405
QY 1452 ----- 1452
Db 61404 GCTGAGTAAATTGAGGTTTCCAGAGGTTCAACGACTCACCCAGAGTCAAGCAAAATGC 61345
QY 1452 ----- 1452
Db 61344 AGTTGTAAACCAAAATTCAGATGTTCTACAGCGCAGCATCCACTGCAACCCACAGCA 61285
QY 1452 ----- 1452
Db 61284 GGTGACCAACAGAGGCCCTCAGTCCCCCAGGGGCCCGAGCTCAGTAGGGAGTTCC 61225
QY 1452 ----- 1452
Db 61224 GTGCCGATGTACGAGCAGGAGGCTGTTCCGGTGAAGAGCCCTGAAAGGCCACTGTCT 61165
QY 1452 ----- 1452
Db 61164 TCCACATGGCAGAGTGGCTCTTGTGAAGGGGAAGAGATGGGAGCCACCCAGGGG 61105
QY 1452 ----- 1452

Db	61104	TGTGGGCTGTAGCGGAGGAGGACTGGGGTGGGTGTCCTCCGGGAGGGGTTCCAGCTTGTAG	61045	ProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuPro	1555
Qy	1452	-----	1452		1555
Db	61044	GAAGTTTGAAGCCAGGAGAGGAGAGCAGAGCAGTAAAGTCCCTGATTGCCAGGGAGG	60985	CCGGGAGAGCCCTTGCCTATGAGGACACAGGGGACACCTTTGCGGGCACCTCCCA	59905
Qy	1452	-----	1452		1573
Db	60984	GTTTGGCTCTCAGCCCTAGGCAATTATGGAGTCTTTGGAAGCATCCACCGCATGACCAA	60925	ArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlySer-	59904
Qy	1452	-----	1452		59904
Db	60924	GACAGGTCCAGATTCTAGAAATATCTTTTGAAAAACAAGGCGAGTTCCCTCTTTACGA	60865	CGAGGTTGCGCCGTGACACCGCGGAGCCACGCGGCGCTGCAGGAGGTGA-GTGGGG	59846
Qy	1452	-----	1452		1573
Db	60864	CAGTAATGAAGACATCCCTAAATAGAGTTTGTGTCTTGCAAGCCCTAAAGTCCACCA	60805	TGTGCATGGCGGTGAGTGGGGTGGGCGCCTGTCTGGAGAAAGCTGTGCTCCTCCCATCCACC	59786
Qy	1452	-----	1452		1573
Db	60804	TATCAAGTGTCTCGAAGCCTGCTGAAAAAGAGGAGGAGGACCTGGTGGCCTGGCTACGAA	60745	ATTAGCTTAGTTTGCACCTGGGATATCTCGCCACCGCTTTCCACCACATCCAAACCAAC	59726
Qy	1452	-----	1452		1573
Db	60744	GGTCCAGTCTGAGCTGAGACCCCTCCCTTCTCACCACCTCTTCTGTCTGGAGGCA	60685	CTGAGGCCCGTGGGCTCTGCCTCCGATTCAAACCTGTCCAACTCCTTGGCCACCTCCC	59666
Qy	1452	-----	1452		1573
Db	60684	ACACCTCAGCCCTGACCTCAGTGTCTCGGGGTGAAGGCTCAGGGCGGTAGTATGG	60625	AGACCACCGTGTCTCACCCTAGCTTCCCGCAGCGCCCTCTCTCTGCTGTGTAATCC	59606
Qy	1452	-----	1452		1573
Db	60624	GTCTGGGTGTGACTTTTCTGTCTGGCATTTGGGTGGGCATGAGGGTCCCAGGCTGAA	60565	ACTCTGCAACAGCTACCCGGATCTTTCTAAAAATGCAAAATCATATATTTCCACTTCCC	59546
Qy	1452	-----	1452		1573
Db	60564	CAGAGGAATGTTTTTACCACATAGAGGTGTTGGGCTCTTTCTCGCAAACTCCAGGA	60505	TGCTTCCATCCTTCTAGCAACTTCACACATTTTGTATGGCTTGGGGCGCTGCCTGTT	59486
Qy	1452	-----	1452		1573
Db	60504	CCATCAGAGAGCCACCACTCGCGGAGGAGAGTGTGACTGTGAACCTTTTACCCTTT	60445	GGGGCCCTGCTGCCTCTCATTCAGCGGATTCCTGCTCCTCCAGCCCCAGCCCCCTG	59426
Qy	1452	-----	1452		1573
Db	60444	CTCAGTCCCCCAGGAGCGTGGGACACAGGCTCAGGCCAGGCTCGCAGGSGCAGTAA	60385	GGCCCTTTTCTCTTTGTTCCCTGGCCATGCTTAGCTCGGTCAATTCAGTATTTGCTGGG	59366
Qy	1452	-----	1452		1573
Db	60384	GTAACAAGTGTCCATCTCAGGGTTAGCAAGCCCTCTGTCTCCTGCTCTGAGGCAT	60325	GGCTTTGTGCTGCTCCTCTCTCTGCTGCCATGTCCCGCCTTCCAGATCTTTACTTA	59306
Qy	1452	-----	1452		1573
Db	60324	CAGATGTCACTTCCATCTTATAGATGACAAACTTTTGGAGGCTCAGAAAGGGGATGAGC	60265	GTGGGTTCTTTCCATCCCTCAGGTCTTTGTTTACATATTAATTCCTTTGGGAGGCTTCT	59246
Qy	1453	-----	1455		1573
Db	60264	TGGTCTGGGTATGGCTGTGSCCAGGCTAGAGCTTACATCCCTCTGCCCCAGGGCAC	60205	AACCAGACCCCTATCTCCAGTTTCATATCACATGCTGTGACATTTTAAAAATTGCTTCCG	59186
Qy	1456	ProLeuLysTyrAspThrGlyAlaSerThrGlySerLysHisAspValArgSer	1475	1573	59126
Db	60204	CCGCTCAAGTACGACACCGGGGCTCCACCACTGGCTCCAAAAAGACACGACGTACGCTC	60145	1573	59066
Qy	1476	LeuIleGlySerProGlyArgThrPheProValHisProLeuAspValMetAlaAsp	1495	1573	59006
Db	60144	CTATCGGACGCCCGCCGAGGCTTCCACCGGTGACCCGCTGGATGTGATGGCCGAC	60085	1573	59006
Qy	1496	AlaArgAlaLeuGluArgAlaCysTyrGluSerLeuLysSerArgProGlyThrAla	1515	1573	59006
Db	60084	GCCCGGCACTGGNACGTGCTGTCTACAGAGAGGCTCTGAGAGCCGCGCAGGACCGCC	60025	1573	59006
Qy	1516	SerSerSerGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLys	1535	1573	59006
Db	60024	AGCAGCTCGGGGGCTCCATTGCGCGGCGCGCCCGCTCATTTGCTGCTGAGCTGGGTAAG	59965	1573	59006

QY	1573	-----	1573	-----	Db	57805	AATACAAAAA	TAGCCGGGTGTGGCGGCGCCTGTAGTCCAGCTACTCGGGAGGCT	57746
Db	58885	GAAAGAAAAATTTCTCTGAGCGTGTTCACCTGTGTAATCTCATTTGTTGCTAGTTTA	58826	-----	QY	1644	-----	-----	1644
QY	1573	-----	1573	-----	Db	57745	GAGCGAGAGAAATGGCGTGAAACCCGGGAGCGGAGGCTTGCAGTGAAGCAAGATCGCGCCA	57686	
Db	58825	TCACCTGTCTCTCGCATTTGAATGTTCAGCTTGTGAGCGCTGGGATTTCTGTTTCGTTCACT	58766	-----	QY	1644	-----	-----	1644
QY	1573	-----	1573	-----	Db	57685	CTGCATCTCAGCCTCGGGCGACAGAGCGAGACTCTCTCTCTAAATAATAAGAAATTAATAA	57626	
Db	58765	GGGGTGACCCCAAGTTCTCAACAATGCTTTGCCACGTAGTAGAGCTGCATCAATATTTT	58706	-----	QY	1644	-----	-----	1644
QY	1573	-----	1573	-----	Db	57625	AAAAAGAAATAAAAAAAGTTGCATCCCTTTTGGAGTGTAAATCTGCATTTGGGATGTCCTAT	57566	
Db	58705	TTAATTGATTGAGTGAGTGAATGATGAAGAATGAATTTTTTAAAAAACTATAACAAAA	58646	-----	QY	1644	-----	-----	1644
QY	1573	-----	1573	-----	Db	57565	GTTTGGGACAACCTTTGATGCAAAAAGCATCTTCGTAGAAGTCACCTCTTGTGTCTGG	57506	
Db	58645	AGCAAAATGAGTCAGTGAGCAAAAAGTGAACCTAAGGCAATGAAGAAATGAAGAGTGAATG	58586	-----	QY	1644	-----	-----	1644
QY	1573	-----	1573	-----	Db	57505	CGTGATGTTTTCTGCTGTGTCGAGCGCTCAGTTCTGGTGTGTGGCGAGCCACACAT	57446	
Db	58585	AAGAGACCTGGTCTCTTGGGATCCGAGGTCCCTATCTCAACAACCTCCCGTAAATGCC	58526	-----	QY	1644	-----	-----	1644
QY	1574	-----	1574	-----	Db	57445	GTAGTGGGAGAGCTGTCGGGTGCAGAAAGTAGGGGCATCCAGACAGGTGGAGCGACA	57386	
Db	58525	AGCCCCAGAGCCCGATGCAATCCACCTTGGCCGTCCACAGGCGCCTTTCTGTCAGCAAG	58466	-----	QY	1645	-----	-----	1645
QY	1579	AlaSerGlnAspArgLysLeuThrSerThrProArgGluLeuAlaLysSerProHisSer	1598	-----	Db	57385	CCATCAGCGCTAGGTATGGCTGCCTCACATGAGCTCCCTCTGCCCGCGAGCGCTGCC	57326	
Db	58465	GCATCCAGAGCCGAAAGCTGAGCTCGACGCGCTCGTGAGATGCCAAGTCCCGCGACAGC	58406	-----	QY	1649	TyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyr	1668	
QY	1599	ThrValProGluHisProHisProIleSerProTyrGluHisLeuArgGlyVal	1618	-----	Db	57325	TACTACTGCCCCGACACCTTGGCCCCCAACCCCACTTACCCGACCTGTACCCACCTAC	57266	
Db	58405	ACCGTGGCCGAGCACCACACCCCATCTCGCCCTATGAGCACCTGCTTCGGGGCGTG	58346	-----	QY	1669	LeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleLeuAsnAsp	1688	
QY	1619	SerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIlePro	1638	-----	Db	57265	CTCATCCGCGGTACCCGACACGCGCGCTGAGAACCGGACAGCATCATCAATGAC	57206	
Db	58345	AGTGGCGTGACCTTGATTCAGGCCACATCCCTCGCCCTTGCACCCCACTTCATACCC	58286	-----	QY	1689	TyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAsp	1708	
QY	1639	ArgGlyIleProLeuAsp-----	1644	-----	Db	57205	TACATCACCTCGCAGCAGATGCACCAACGCGCGCCACCGCCATGCGCCACGAGCTGAT	57146	
Db	58285	CGCGGCATCCCTCTGAGCGAGGTGATGTCCTGGGCTCCCGAACCTTGACGTGTGC	58226	-----	QY	1709	MetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyPro	1728	
QY	1644	-----	1644	-----	Db	57145	ATCTGAGGGGCTCTCGCCCCCGAGTCTCTCGCTGGCACTCAACTACGCTGGGGTCCC	57086	
Db	58225	TGAACAGGGCCACGGACCTCATCAGTGTTCGCTCAGGGAATCCTTAGGCATCAACTGTCA	58166	-----	QY	1729	Arg-----	-----	1729
QY	1644	-----	1644	-----	Db	57085	CGAGGTGAGTGGGTGGCGACACCCCTCCGCTGGGTTTGGCCTTATTTCCCAAGGACATG	57026	
Db	58165	GGTTCCCTCGATGGCGAAACTGAGGCTCGGGATTTGGAAGACCAACAGTGAATCATG	58106	-----	QY	1729	-----	-----	1729
QY	1644	-----	1644	-----	Db	57025	GGCGTCCCTGTGTGGCTCGCGAGGACGCTAGACCTGTGTACCTTGTGGGTCACTTGT	56966	
Db	58105	AGCTTAGGTTGGAGCAGAAATTTCTCTTAGTAGTTTGCAGGACATGTGGGGTTAAACATTT	58046	-----	QY	1729	-----	-----	1729
QY	1644	-----	1644	-----	Db	56965	GTGAACGACCTCAGTGGGTGGGCTGGGGTTGTGCGTGTGTGGTCTCTGTTGGCATCT	56906	
Db	58045	CAGTGGTTTTCTTTCCGGCAGGACTTATCAGTGCCTTTTAGCAATGCAAGGTATAGAAT	57986	-----	QY	1729	-----	-----	1729
QY	1644	-----	1644	-----	Db	56905	GGTAGGTGAGTGACAGCGTGTGGCTCTGGCTGCATCCTCAGTGGGTGTGCGTGCATCT	56846	
Db	57985	GAGGACTTGAGTATATGCAATTTTCAAAATAGACATGATCTGAAAGTCTTTTAAAGTT	57926	-----	QY	1729	-----	-----	1729
QY	1644	-----	1644	-----	Db	56845	GTGTATACTCTTAGGATACAGGGGCTCAGAGGTTTAAAGATCAAAATGTGGCGGGCAC	56786	
Db	57925	GCCGGCAGCGTGTGTACACCTGTAATCCAGCACTTTTGGAGCGCGAGCGCGGAT	57866	-----	QY	1729	-----	-----	1729
QY	1644	-----	1644	-----	Db	56785	AGTGGCTCATGCTGTAAATCCAGCACTTGGGAGGGCGGAGGAGGTGGATACAAAGTTC	56726	
Db	57865	CACAAGTCAAGGATAGAGACCATCTCGGCTAACCGGTGAACCCCGTCCCTACTATAA	57806	-----	QY	1729	-----	-----	1729
QY	1644	-----	1644	-----					

Db	52348	CCTAACCTCATGAGCCCGCTCTTGCTGCCCAAGAGAGCGCCCCCGGGTGC	CGCCGACAGAG	52289
Qy	1950	ArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeu		1969
Db	52288	CGGCCCGAGCAGACACCGGCCATGCTCTCTCCCAAGCCCCCAGCCGCTCCGGGCTG		52229
Qy	1970	GluProAlaSerSerProSerLysGlySerGluProArgProLeuValProProValSer		1989
Db	52228	GAGCCCGCTCTCTCCCGCAGCAAGGGCTCGAGCGCCCGGCCCTAGTGCCTCTCTCT		52169
Qy	1990	GlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerPro		2009
Db	52168	GGCCACGCCACCATCGCCCGCACCCCTCGCAAGAACCTCGCACCTCACCGCCGCG		52109
Qy	2010	AspProProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLys		2029
Db	52108	GACCCGCGGGCCACCTGCTCGCTCGGCTCGGACCGGCACCGGGAAGACTCAAGTAA		52049
Qy	2030	ProPheSerIleGlnGluLeuArgSer		2040
Db	52048	CCCTTTTCCATCCAGGAACCTGAACTCGTCTCTGGGTAAACACCCCTGACAGCGCC		51989
Qy	2040	-----		2040
Db	51988	ACCTTCATAGACGCGATTATCATCGTCAAAATGCTCACGNATAAAGGGCGCAGAA		51929
Qy	2040	-----		2040
Db	51928	GGTGGCTGGCCACGGCTCCCTCGCGATGGTAAGACTTCGGCGCCGCACCCACCCGT		51869
Qy	2040	-----		2040
Db	51868	CTCGTGTCCAAAGATATTTTCAGATCTCTGCTTTTACTTTTGGCCCCGTTTTTTGT		51809
Qy	2040	-----		2040
Db	51808	TGTTGGTTTTGGTATTTTGTGTAAGCCCATCCGCTCTCGCCGTTTTTGACGCGTGAC		51749
Qy	2040	-----		2040
Db	51748	GACTACTCCGGCCCGGCTGCCCCCTTGTTGGGGGGGCTCATCTTTGCACATCATTT		51689
Qy	2040	-----		2040
Db	51688	TACCATGTTTTTTTTTTTGGATTTTGTGCTTTTCTTTTCTTAATGAATGATCTGTGAT		51629
Qy	2040	-----		2040
Db	51628	TCTGACTTCGACTCGGCCCCCATCTCCCTCTTTTGGCCTGTGTCCAGGGAGCAGGATGG		51569
Qy	2040	-----		2040
Db	51568	GGCTGGGGAGGGCTCGGGCCCTACGCCCCACCTGCGCGGTGCCTGGATGCTGCGACT		51509
Qy	2040	-----		2040
Db	51508	GGGGGAAGTGGAGCAGCGGTGCAAGGAAGCTGAGGGGGGCGAGGACCTCGCGTGT		51449
Qy	2040	-----		2040
Db	51448	CGAGGAGGAGTGGGTCTGGCTTTGATCTTGCCCTGTCCCGAGCCCTGTATCCCCAGA		51389
Qy	2040	-----		2040
Db	51388	AAAAGGGAGCCCTCTGCTCTGAGACCCCTGCTTGGCCCTAGTTATGGCTCTCTCTG		51329
Qy	2040	-----		2040
Db	51328	TTGGACTGGGATGSCCGAGGCTATAGCCAGCGGGGCCCCGGGACCCAGGGTCACTCC		51269
Qy	2040	-----		2040
Db	51268	CAGCCACACCCCCACTTCTCACTCCGCCCCACACACTCTCTTCCCCAGAGACCCATGCTGC		51209

Qy	2040	-----		2040
Db	51208	CCCCATCTACGCTGGCCTCGCCCGGCTCCACACCTGAACCCATCTCTGTCCCTTCTT		51149
Qy	2040	-----		2040
Db	51148	GCCTAATCTCTCTGTGTCTCTCCCTCTCTGTGTCTCTGTGTCCCGGGCTCTGCATCTC		51089
Qy	2040	-----		2040
Db	51088	TCCACCTCTCCCTTGGCCTCCCTGTCTCTCCCCAACACCCCTCTCTGCTTACTGTCTTT		51029
Qy	2040	-----		2040
Db	51028	GGGAGCCCAAAACCTACCCCTAGCTTGGGTTCCCTTTGACCCCCCGGGTCCACGCCAG		50969
Qy	2040	-----		2040
Db	50968	CTGGAGGGCAGCCCTCGCCCTCGGGCTCGGAAACCTGGGCCCGGTCCTGACTCTGCA		50909
Qy	2041	-----LeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProVa		2056
Db	50908	CCCCCGCCTGCCCCCTAGGTTACCACGCGCAGAGCTACAGCCCCGGAAGGGTGGAGCCGT		50849
Qy	2056	lSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluG1		2076
Db	50848	CAGCCCTGTGAGCTCACCCAGTCTGACCCACGACAAGGGCTCCCCAAGCACCTGGGA		50789
Qy	2076	uLeuAspLysSerHisLeuGluGlyLeuArgProLysGlnPro		2091
Db	50788	GCTCGACAAGAGCCACCTGGAGGGGAGCTGCGGCCCAAGCAGCAGGTACGCCCCACCC		50729
Qy	2091	-----		2091
Db	50728	AGTACCCAGGCCCCGGAAGCCCTGCACAGTAGAGACCCCTCAAGGCCCATCATGCAGATA		50669
Qy	2091	-----		2091
Db	50668	GGAAACAGAGGTGCCTAAAGGCCAAGGAATTGGCTGGATCATGAGGCTCAAAGGCATGG		50609
Qy	2091	-----		2091
Db	50608	GGCTGGGATTGGGCCCCAGCAGCCCTGGGGCCAGACAGACCCCTCACCAGCTCTGTCTC		50549
Qy	2092	-----GlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProL		2109
Db	50548	TGCCTGCAGCCCGCTGAAGCTTGGCGGGAGGCGCCCACTCCACACCTCGCGGCGC		50489
Qy	2109	euProGluSerGlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGlyH		2129
Db	50488	TGCTGAGAGCCAGCCCTCTGTCAGCCCTGTCTCCAGACGCGCCCGAGGGGTCAAAGGTC		50429
Qy	2129	isGlnArgValThrLeuAlaGlnHisIleSer		2140
Db	50428	ACCAGCGGTGGTCACCTGGCCCGCAGCACATCATGTAACCTACGCGTTCTCTGTCTGTGC		50369
Qy	2140	-----		2140
Db	50368	TTGTACCTTTTGACCTGGGGGCACAGGCTTGGAGAGGGGATGGGGAACCCCAAGCCC		50309
Qy	2140	-----		2140
Db	50308	TTCTGTCTGGCGGGTGGCTGGGGATCCAGGGCATGGCGCTGGGGGATCCAGGGCGT		50249
Qy	2140	-----		2140
Db	50248	GGGTAGGGGTAGATCCAAAAGCCCGAGCACCGGCACCATCACCGCCCCCTAATCCATG		50189
Qy	2140	-----		2140
Db	50188	GGAGGAGCCTGTGATGCGAGCCGATGGCATTTTCAOGGGCAATGAGGCCCTTCTGTGTGC		50129

QY	2140	-----	2140	Db	49048	ACTCCCACTCCACCCCGCCGCCACCCCACTGGTGAACCGTGGGCGCTCTGGGGT	48989
Db	50128	CCAGGTTTCTCAGTGTCTATGGGCTGGTCTCATCAGCCATCTGCCAACTACCAGCTTGGGA	50069	QY	2204	-----	2204
QY	2140	-----	2140	Db	48988	CAAAAGAGAAAGAGAGGGAGGGCCCTGAGCTCCAGGGTGAAGAGGCGTCTTGGTGGGG	48929
Db	50068	CCGTGACCAACAGCCCACTCCCATGACACTGGGACACGAGGCCCCAGAGGGTGGCGGG	50009	QY	2204	-----	2204
QY	2140	-----	2140	Db	48928	GCGGGGGGTAATGTGTGTGCGAGACAGACATGSCAAGACAGCAGGACATCTTTGGGG	48869
Db	50008	CAGGTCCACAGTCACCCAGGAAGCTGGCCCCCACCAGGATTCGCCCCGAGCTCCGTCTA	49949	QY	2204	-----	2204
QY	2140	-----	2140	Db	48868	GGCAGTGGTGTAGCTGGCACTGGGTACAACACAGAAATTCAGAGCAGGGTCAATAAACATG	48809
Db	49948	GCCCCCTCCCAACCCCAAGAGTTCTGTCAAGAGAGTGTGCTGACTCTGGSCCCCCC	49889	QY	2204	-----	2204
QY	2141	-----	2141	Db	48808	TGCCCCATGGCCAGATCTAGCCCCGGGCCCTCTGTTTGTACAATTCATGAACATAAAAAA	48749
Db	49888	ACTTGCCTGCAGGAGGTCAACACAGAGACTACACCCGGCACCAACACAGCAGCTCAGC	49829	QY	2204	-----	2204
QY	2157	AlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeu	2176	Db	48748	TGATTTTACATTTTAAAGGTTGTTTAAAAAATAAATAATGATAACGATACATGCC	48689
Db	49828	GCACCCCTGCCCGCCCCCTCTACTCTCCCTGGGGCCAGCTGCCCGCTCTGGACCTC	49769	QY	2204	-----	2204
QY	2177	ArgArgProProSerAspLeuTyrLeuProProAspHisGlyAlaProAlaArgGly	2196	Db	48688	AGAGATTACTTGTGGTTCGAAAAATGCCTAAACCGTTTATGATTTGGCCACGTACAGGAAA	48629
Db	49768	CGCGCCCAACCACTGACCTTACCTCCCGCCCCCGGACCATGTTGCCCGGCCCGTGGC	49709	QY	2204	-----	2204
QY	2197	SerProHisSerGluGlyGlyLys-	2204	Db	48628	GCGTGTGGGCCCCGTGTTTAGAGTGAGGAGGGTGCGCCCTCAGTAGGGAGACCTCT	48569
Db	49708	TCCCCCACACGCAAGGGGCAAGAGGTGAGCGAGGATGGGGTGCCTCTTGCTGTGTC	49649	QY	2205	-----	2215
QY	2204	-----	2204	Db	48568	GACCACATCTGGGGCCCTTTCTCCATCCAGGTCTCCAGAGCCAAACAGACGTCTGTC	48509
Db	49648	GCGGGGGAGGCTTCGTGAGTGCCTGCCTCAGAGCTGGTCTGAGCTCAGCTGCATG	49589	QY	2215	uGlyGlyGlyGluAspGlyIleLeuProValSerProGluGlyMetThrGluProG	2235
QY	2204	-----	2204	Db	48508	GGGTGTGTGGAGACGGTATTGAACCTGTGTCTCCACCGAGGGCATGACGGAGCAGG	48449
Db	49588	TGAGGGCCGCTGGTCCATTGCACAGGCAAGAAATGGAGGCTCCAGGAGATGGGGAGT	49529	QY	2235	yHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSe	2255
QY	2204	-----	2204	Db	48448	GCATCCCGAGTGTGTGTATACCCGCTGTGTACCCGGATGGGGAAACAGACGAGCCAG	48389
Db	49528	CACAGACGGATACATAACGAATAATAGCTGATGAAGTACATGCTGTGCTCAGTGGTT	49469	QY	2255	-----	2255
QY	2204	-----	2204	Db	48388	-GTACTTCTGTGGGCACATGCGTGCCTCCCGGGATGCTCTGGTATCCCTTGCCCATCCTT	48330
Db	49468	CCAGGCACTCAGTGTACCTTTACACACAGCCCCCAGAGACATAGGCAGCATTCCTC	49409	QY	2255	-----	2255
QY	2204	-----	2204	Db	48329	GGCCCCAGTCCACCGTGGTGCCATGTGGAGAGTGACAGGGCACAGGGCTCAGTGGGTG	48270
Db	49408	CCATTTACAGATGAAGAACTAAGGCTCGAGAGTGAATTCATTTGCCCAACAGGGGCA	49349	QY	2255	-----	2255
QY	2204	-----	2204	Db	48269	ACCTCAAGCTGCAAGCAGGTTTACCAACTTTGGGGGTGTGATACATGCCCAACCTCC	48210
Db	49348	CAGTCAGTAGGTGACAGGGGGATTGTTCCAGGGCTGTATGATTAGAGAGCAGCCTTT	49289	QY	2255	-----	2255
QY	2204	-----	2204	Db	48209	TGGGAGACCTCAGTGTAGTCTTTCCATGCTTTGGGCTGGGATCTCAGCTGGACACAGGCC	48150
Db	49288	CCACATGCACACCCACACACATGCGCACACACATGCAACCCGACACATGCGC	49229	QY	2255	-----	2255
QY	2204	-----	2204	Db	48149	CCGTTTGACAGCCCTGTTGAACCTCTAAGAAATAATGAGTAGTGTGTGCGGCCAGAC	48090
Db	49228	ACCGGCACACACATGTCACACCCACACATATGCGCGGCACACATGCACACCCGC	49169	QY	2255	-----	2255
QY	2204	-----	2204	Db	48089	CGGGGGCGATGGGTGAGGTGGGACCTGAGAAGAAAGCTGGGCCCTGCCCCCTGGGAAG	48030
Db	49168	ACATGCGCGGCACACACATGACACACCGGCACACATGCGGCACACACATGCACAC	49109	QY	2255	-----	2255
QY	2204	-----	2204	Db	48029	GGCATGATCGGAACCAATTTAGTCTTGGGGCTCTCTTGAGAGGGTCAGGCTGGAGCA	47970
Db	49108	CCGCACACAGCGGCACCCGCACACCCCTGTAGACACAGGCACATGCATGCTCAGCAC	49049	QY	2255	-----	2255
QY	2204	-----	2204				

Db 47969 AGTGGTCAGAGGAGCCACACGACGACGAGCGTCTCCCAAGACACCCCTAGCCCCAGG 47910
Qy 2255 ----- 2255
Db 47909 ACGGGTGAACCGCAGAGGAGATTTTTCAGGAGCGGTGATCTTCTACAGGAGGGGATGCAG 47850
Qy 2255 ----- 2255
Db 47849 GCGTGGGGGGGTGAAGCTTGTCTTCCAAATGTCTAAGGATCTCTCAGGTGGTGGTTCCTCC 47790
Qy 2255 ----- 2255
Db 47789 CATCATCAAGGCGATGCAAGCTGGCCACCAAGTGAAGTGTGAGTGTGATGCTTGTATCTCA 47730
Qy 2256 ----ArgMetGlySerLysSerProGlyAsnThrSerGlnProAlaPheSerLy 2274
Db 47729 CAGCAGATGGCTCCAGTCTCAGGCAACACAGCAGCGCCAGCCTTCTTCAGCAA 47670
Qy 2274 sLeuThrGluSerAsnSerAlaMetValLysSerLysGlnGluLeuLeuLeuLysLe 2294
Db 47669 GCTGACCGAGAGCAACTCCGCGCATGTCAAGTCCAGAGCAAGAGATCAACAAGAGCT 47610
Qy 2294 uAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluLlePh 2314
Db 47609 GAACACCCACACGCGAATGAGCTGAATACAGTAAGGCGCTGCAGGCTCCCGGGGAAG 47550
Qy 2314 eAsnMetProAlaIleThrGly 2321
Db 47549 CATGGGGCCACAGTGGGGCGG 47528
RESULT 21
AF303586
LOCUS AF303586 3120 bp mRNA linear PRI 01-FEB-2003
DEFINITION Homo sapiens NCOR isoform b mRNA, complete cds.
ACCESSION AF303586
VERSION AF303586.1 GI:28190007
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3120)
Yu, L.
Direct Submission
Submitted (08-SEP-2000) Institute of Genetics, Fudan University,
No. 220 Handan Road, Shanghai 200433, P. R. China
FEATURES
source Location/Qualifiers
1..3120
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
85..2829
/codon_start=1
/product="NCOR isoform b"
/protein_id="AA032942.1"
/db_xref="GI:28190008"
/translations="MSSVYPPNQGAFSTQSRSPHVSVKYTFPSTHHOODPAFGKH
EAPSPITSGQPCDDONASPSKLELIQMDRVDREIAKVQOILKLKKQOQLEE
EAAKPPEKPPVPVQKRSIVQIYIDENRKKAAEAKHIFEGLPKVELFLYNQ
SDTKVHENITQVMKKLILFFKRNHARKOREQIKORYDOLMEAWKKVDRIEN
NPRKAKESKTRYEYKQFEIRKOREQORFORVGORGAGLSATIAARSEHSEIID
GLSEQSPSKQMKQLSVIPPMFDAQRRVKFINMGLMEDPMKVYKDRQFMNWTDDH
EKEIFDKFIQHPNFKGLIASYLERKSVPCVLYLTTKNENKALVRNRYKRRGR
NQOIAEPSQERKEEEDKAEKTEKKEEKDEEKDEKSKENTKEKDKIDGTAE
ETEPERQATPRGRKANSQGRKGRITRSMTNEAASASAAAAAETPEPPPLPPPEP
ISTEPVETSTEEEMVAKGLVHEHWNAAIAKVMGTSEAKQCNFYNYKRRNL
DNLLQHQKTSKREEDVQCSVASTVSAQEDIEDASNEENPEDESEAGNS
DTESAPKSPKSPEDSPENASTVPAVELEPTTETAPSPSPSLAVSPSPA
EDESVEVQNDSTISAETAEQMDVDQOEHSAEESVCDPPPATKADSVDEVPEPHA
SKYEGNTKREDLDRASEKVEPRDELVVAQINAOBPPEQSDNDSSATCSADEVDG
EPRQRMPMDSPSLNPTGSLVSPSLKPNFLDLPQLQHRRAVTPMWSCTPCNP
IGTPVSGYALYQRHIAKHMSALLEBQORQEQIIDLCSSTSPCGTSKSPNREWSR

polyA_signal 3087..3092
ORIGIN
Alignment Scores:
Pred. No.: 1,05e-36 Length: 3120
Score: 2187.00 Matches: 500
Percent Similarity: 61.44% Conservative: 115
Best Local Similarity: 49.95% Mismatches: 216
Query Match: 16.55% Indels: 170
DB: 9 Gaps: 30
US-09-522-753-5 (1-2517) x AF303586 (1-3120)
Qy 137 LysAspArgSerLeuThrGlyLysLeuGlu---ProValSerProProSerProHis 155
Db 190 CAGGATCCAGCATTCGGAGGCAAAACATGAAGCTCCATCTCTCCAAATTCGGGGCAACCA 249
Qy 156 ThrAspProGluLeuGluLeuValProProArgLeuSerLysGluGluLeuLeuGlnAsn 175
Db 250 TGTGGAGATGATCAAAATGCTTCACCTTCAAACTCTCAAGGAGAGATTAAATACAGAGT 309
Qy 176 MetAspArgValAspArgGluLeuThrMetValGluGlnGlnIleSerLysLeuLysLys 195
Db 310 ATGGATCGTGTAGATCGAGAAATTCGAAAAGTAGAACAGCAGATCCTTTAACTGAAAGAG 369
Qy 196 LysGlnGlnGlnLeuGluGluAlaAlaLysProProGluProGluLysProValSer 215
Db 370 AAACAACAACAGCTTGAAGAGAGGAGGAGCTAAACCTCTCAGCTGAGAGAGCCGCTGCC 429
Qy 216 ProProProLeuGluSerLysHisArgSerLeuValGlnIleLeuLysAspGluAsnArg 235
Db 430 CCTCTCTCTGGAGCAGAAACACCGCAGTATTGTGCAAAATTTATTTATGATGAGAAATCGG 489
Qy 236 LysLysAlaGluAlaAlaHisArgLysLeuGluGluGluProGlnValGluLeuPro 255
Db 490 AAAAAAGCAGAGAGAGCTCATAAAATTTTAAAGTCTTGCCCAAAAGTTGAATGCCA 549
Qy 256 LeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAla 275
Db 550 CTGTATAACCCAGCATCAGATACCAAGGTGTACCATGAGAACATCAAGACAAACAGGTG 609
Qy 276 MetArgLysLysLeuLeuLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLys 295
Db 610 ATGAGGAAAAAATCATATTTTATTTTAAAGAGAAATCATGCAAGAAAAACAAAGGAA 669
Qy 296 GlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArg 315
Db 670 CAAAAAATCTGCCAGCGCTTATGATCAGCTCATGAGGATGGGAGAAAAAGTGGACAGA 729
Qy 316 IleGluAsnAsnProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLys 335
Db 730 ATAGAAAATAATCTCGAGGAAAGCTAAAGAAAGCAAAACAGGGAATACTATGAAAAG 789
Qy 336 GlnPheProGluLeuArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGly 355
Db 790 CAGTTTCCAGAAATTCGAAACAAAGAGAAACAGCAAGAAAGATTTCAG---CGAGTTGGG 846
Qy 356 GlnArgGlySerGlyLeuSerMetSerAlaArgSerGluHisGluValSerGluLeu 375
Db 847 CAGAGGGAGCTGGTCTTTCAGCCACCATTCGTAGGAGTGAGCATGAGATTCTCGAAAT 906
Qy 376 IleAspGlyLeuSerGluGlnGluAsnLeuLysGlnMetArgGlnLeuAlaValIle 395
Db 907 ATTGATGGGCTCTCTGAGCAGGAAAGTAATGAGAAACAATCGGCAGCTCTCTGTGATT 966
Qy 396 ProProMetLeuTyrAspAlaAspGlnGlnArgLysPheLeuAsnMetAsnGlyLeu 415
Db 967 CCACCTATGATGTTGATGTCAGAACAAAGCAGTCAAGTTTCATTAAATCATGAATGGGCTT 1026
Qy 416 MetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGln 435
Db 1027 ATGGAGACCCCTATGAAAGTGTATAAGATGAGGAGCTTTTATGAATGTTTGGACGTGACCAT 1086

LOCUS BC054296 1917 bp mRNA linear VRT 03-JUL-2003
 DEFINITION Xenopus laevis cDNA clone IMAGE:6879542, partial cds.
 ACCESSION BC054296
 VERSION BC054296.1 GI:32450290
 KEYWORDS
 SOURCE Xenopus laevis (African clawed frog)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 Xenopodinae; Xenopus.
 1 (bases 1 to 1917)
 Klein, S.L., Strausberg, R.L., Wagner, L., Pontius, J., Clifton, S.W.
 and Richardson, P.
 TITLE Genomic and genomic tools for Xenopus research: The NIH Xenopus
 Initiative
 Dev. Dyn. 225 (4), 384-391 (2002)
 JOURNAL 22341132
 MEDLINE 12454917
 PUBMED
 REFERENCE 2 (bases 1 to 1917)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Vallalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
 TITLE Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 JOURNAL 22388257
 MEDLINE 12477932
 PUBMED
 REFERENCE 3 (bases 1 to 1917)
 Klein, S. and Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (30-JUN-2003) National Institutes of Health, Xenopus Gene
 Collection (XGC), National Institute of Child Health and Human
 Development, 6100 Executive Boulevard, Room 4801, Rockville, MD
 20892-7510, USA
 NIH-MGC Project
 CONTACT: XGC help desk
 Email: cgabs-x@mail.nih.gov
 Tissue Procurement: Drs. Donald Brown and Liqian Cai
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada
 info@cgsc.bc.ca
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
 Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
 Ness, Pawan Pandoh, Anna-Lilias Prabhu, Parvaneh Saedi, Jacqueline
 Schein, Duane Smalios, Michael Smith, Lorraine Spence, Jeff Scott,
 Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
 George Yang, Scott Zuyderduyn, Marco Marra.
 Clome distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 49 Row: 1 Column: 16.
 Location/Qualifiers
 1..1917
 /organism="Xenopus laevis"

/mol_type="mRNA"
 /db_xref="IMAGE:6879542"
 /clone="IMAGE:6879542"
 /tissue_type="Whole, tadpole, metamorphosis, stage 53"
 /clone_lib="NICHDP_XGC_Tadl"
 /lab_host="DH10B"
 /notes="Vector: pDNR-LIB"
 337..>1917
 /codon_start=1
 /product="Unknown (protein for IMAGE:6879542)"
 /protein_id="AAH54296.1"
 /db_xref="GI:32450291"
 /translation="MSGTQQAQTQSRRTGEARYPPHAPMAYPMQLSRAHTDANLEYOH
 HTDYSHLGHGSLVLTQRRPSSLFQFQGNRSQELHGRDQHSYHLHWSKQHDLD
 LABAKPRLDILHPLMRSLITQNOHAIGEELNKKDKRMTGKLEPSPSPPLADGDL
 DLLPRLSLKEELIMDRVDREIIMVEQOITLKKKQOQLEEAARLEPEKPSPP
 VESKESLQIIVDENRKAQAARILLEGIPHEVLPYLNQPSDTTKOYHENIKINQTM
 RKLLILYFRNRHARKQEQKLCQRDQLMEAWKKVERIENPRRAKSKIREYE
 KQPEIRKQRELOQRMQRVGQNGSLNSAARSEHEVSEIIDLQSQENLERQOQL
 AVTPMLFAEQORIKFINMGLMDPMKYKROVMNWSIQEKDFTREKFIQOPKN
 FVPIASFLERKTVSDCVLFYLTQKNENYKSLIRNRYRRRGKQKQMPRTNQEKDEKE
 KEVDKEYKQKKKKKKKK"

ORIGIN

Alignment Scores:
 Pred. No.: 3,14e-36 Length: 1917
 Score: 2155.00 Matches: 414
 Percent Similarity: 86.58% Conservative: 57
 Best Local Similarity: 76.10% Mismatches: 55
 Query Match: 18
 DB: 5 Gaps: 3

US-09-522-753-5 (1-2517) x BC054296 (1-1917)

Qy 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20
 Db 337 ATGCTGGATCTACTCAGCAAGCCACACAGCGGACGAGGAGCGGCGGTACCTCCT 396
 Qy 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
 Db 397 CCCATGCTATGGCTTATCCATGTCAGTTGTCTGTCGCGCACACGAGTGCACCTCTTG 456
 Qy 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
 Db 457 GAGTACCAGCACACACCGCGGATTACAGCTCCACCTAGGCGCACGCTCTGTTCTTTTG 516
 Qy 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAenGluArgSerGln 80
 Db 517 ACAAGAAGAAGAGACCCCTCCTGCTATCTGAATTCAGCCAGGAATGAAGGTCTCAA 576
 Qy 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSer---Glu 99
 Db 577 GAGTTACATGGAGAAGAGACCAACATCTCTACTGTCATGATGACATGACCAACACATGAT 636
 Qy 100 MetGluPheIleGluSerTyrArgProArgLeuGluLeuLeuProAspProLeuLeuArg 119
 Db 637 CTAGACTTGGCAGAGGCAAGAGCGCTAGATTGGATTCTTCATGACCCACTTATGCGA 696
 Qy 120 ProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArg 139
 Db 697 CACTCA---CTTATAACGCAAAACACCATGCAATAGGGAAGAAATTAACAGAGATCGC 753
 Qy 140 SerLeuThrGlyLysLeuGluProValSerProProSerProHisThrAspProGlu 159
 Db 754 TGCATGACTGGGAAGCTAGAGCTGTATCCCTCCAGTCTCTCTCTAGCTGATGGAGAT 813
 Qy 160 LeuGluLeuValProProArgLeuSerLysGluLeuIleGlnAenMetAspArgVal 179
 Db 814 CTGGATCTTTTGGCTACCAGGCTTTCCACAGGAAGAGCTGATCAGAAATATGGACAGATA 873
 Qy 180 AspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGln 199
 Db 874 GATCGGAATATCATCTGTTAGAGCAACAGATCACCAAGCTGAAGAAAGAGCAACAG 933


```
RESULT 23
AB093281
LOCUS      6328 bp      mRNA      linear      ROD 07-FEB-2003
DEFINITION Mus musculus mRNA for mKIAA1047 protein.
ACCESSION  AB093281
VERSION     AB093281.1  GI:26006244
KEYWORDS    FLI CDNA.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS     Okazaki,N., Kikuno,R., Ohara,R., Inamoto,S., Hara,Y., Nagase,T.,
            Ohara,O. and Koga,H.
TITLE       Prediction of the coding sequences of mouse homologues of KIAA
            gene: 1. The complete nucleotide sequences of 100 mouse
            KIAA-homologous cDNAs identified by screening of terminal sequences
            of cDNA clones sampled from size-fractionated libraries
            Unpublished
JOURNAL     2 (bases 1 to 6328)
AUTHORS     Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
TITLE       Direct Submission
JOURNAL     Submitted (04-OCT-2002) Hisashi Koga, Kazusa DNA Research
            Institute, Laboratory for Genome Informatics, 2-6-7
            Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
            (E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
COMMENT     The CREATE program supported by Japan science and technology
            corporation; cDNA full insert sequencing; Kazusa DNA Research
            Institute; cDNA library construction, clone selection and 5'- &
            3'-end one pass sequencing.
FEATURES
            1. .6328
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            /clone="mbg00559"
            /tissue_type="brain"
            /dev_stage="adult"
            /note="vector:modified pBC SK+"
            1. .6328
            /gene="mKIAA1047"
            <1. .5175
            /gene="mKIAA1047"
            /note="CDS is predicted by in silico analysis. Start codon
            is not identified."
            /evidence=not_experimental
            /product="mKIAA1047 protein"
            /protein_id="BAC41465.1"
            /db_xref="GI:26006245"
            /translation="KASRKRPERDVSQCVASTVSAQEDIEDIASNEENPEPSEGE
            AENSSTESAPSPVPEAAKSSSEDSSENAASRGNTPEVALEATTDPAPCASPSSAVP
            TTKPAERSEVAQVTDSDASAEATPEMDVDHECGAGSSVLPDPATPKADSDPEMQV
            PENTAKSGDADAKERLESTSEKTEARDEDVVAEGIERPEQSDDDSATCSADSGV
            DEPERQRFVPMADKPSLLTPPGSILISSPIKPNLDLQLQRAAVIPWVSCFPCN
            IPIGTVSGYALYQRHAKHESALLERQEQEVDLECRSTSCSTKSPKRWRE
            VLQAPHOVITLNPGLPRTPTPPPLIPSSKTVVASEKPSIFMGSSISQGTFG
            TVLSSHOAVPOEAPKPSVGSISLGLPQOESTKAAPLYIKQEEFSPSONSQPGL
            LVRAQHEGVRTAGAVOEGSIIRGTTPASKIVETISLRSSTTQGTPLPAQAGIPTE
            ALVKGPVSNPIEESPEKVRSEAKSHVIEGKSGHTLSYDLSLCRALPRGPHSD
            LKERTVLSIMQGTPTAESEFDGLKPKQIKRESPIRAPEGAITKGPYDGIIT
            IKEMGSIHEIPRODILTQESRKTPEVQSTPIEGSISQGTPIKFDNNSSQSAIKH
            NVKSLITGSPKLMLEIVPENIKVVERGTDVKAGEPVRAHRTSVVSSSGSVLRS
            TLHEAPKALSPGLYDSDSARTPVSYQNTISRGSPMMNRTSDVSSKASHERKSTL
            TTPQRESIPAKSPVGVDPVSHSPDPHRSAGEVVRSHLPTHLDPAMPFRHALD
            PAAAILQRQLSPTEGYSQQLYAMENTROTILNDYITSOQMVNLAPDVTGRLSP
            REQPLGLPVAIRGIIIDLNTMPTLIVPHAGGSTPTPMDRITVITGTVTFPPPYNA
            ASLSPGHPTHLAASAERERERERERERERERERERERERERERERERERERERER
            QPGRGSHGVYRSPSPVSTQETILQORPSPQGTNGTSTVITLPTAQLRMLPSG
            GPSISQGLPASRVNTAADAALVDAAPQMDVSKTKESKHEAARLEENLRSGAA
            VSEQQOLEOKNLVEKRSVCQCTSSALSGKAQPHASVYSEAGKDGPPKPSRYEE
            ELTRGKTTITANFIDVILITQIASDKDAREGSSDSSSSLSHRYETASDAIEV
            ISPASSFAPPQKFPQAIQFDMVKANAENESTRQYEGPLHHYRSQSESPPQPPLP
```

```
AB093281
LOCUS      6328 bp      mRNA      linear      ROD 07-FEB-2003
DEFINITION Mus musculus mRNA for mKIAA1047 protein.
ACCESSION  AB093281
VERSION     AB093281.1  GI:26006244
KEYWORDS    FLI CDNA.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS     Okazaki,N., Kikuno,R., Ohara,R., Inamoto,S., Hara,Y., Nagase,T.,
            Ohara,O. and Koga,H.
TITLE       Prediction of the coding sequences of mouse homologues of KIAA
            gene: 1. The complete nucleotide sequences of 100 mouse
            KIAA-homologous cDNAs identified by screening of terminal sequences
            of cDNA clones sampled from size-fractionated libraries
            Unpublished
JOURNAL     2 (bases 1 to 6328)
AUTHORS     Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
TITLE       Direct Submission
JOURNAL     Submitted (04-OCT-2002) Hisashi Koga, Kazusa DNA Research
            Institute, Laboratory for Genome Informatics, 2-6-7
            Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
            (E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
COMMENT     The CREATE program supported by Japan science and technology
            corporation; cDNA full insert sequencing; Kazusa DNA Research
            Institute; cDNA library construction, clone selection and 5'- &
            3'-end one pass sequencing.
FEATURES
            1. .6328
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            /clone="mbg00559"
            /tissue_type="brain"
            /dev_stage="adult"
            /note="vector:modified pBC SK+"
            1. .6328
            /gene="mKIAA1047"
            <1. .5175
            /gene="mKIAA1047"
            /note="CDS is predicted by in silico analysis. Start codon
            is not identified."
            /evidence=not_experimental
            /product="mKIAA1047 protein"
            /protein_id="BAC41465.1"
            /db_xref="GI:26006245"
            /translation="KASRKRPERDVSQCVASTVSAQEDIEDIASNEENPEPSEGE
            AENSSTESAPSPVPEAAKSSSEDSSENAASRGNTPEVALEATTDPAPCASPSSAVP
            TTKPAERSEVAQVTDSDASAEATPEMDVDHECGAGSSVLPDPATPKADSDPEMQV
            PENTAKSGDADAKERLESTSEKTEARDEDVVAEGIERPEQSDDDSATCSADSGV
            DEPERQRFVPMADKPSLLTPPGSILISSPIKPNLDLQLQRAAVIPWVSCFPCN
            IPIGTVSGYALYQRHAKHESALLERQEQEVDLECRSTSCSTKSPKRWRE
            VLQAPHOVITLNPGLPRTPTPPPLIPSSKTVVASEKPSIFMGSSISQGTFG
            TVLSSHOAVPOEAPKPSVGSISLGLPQOESTKAAPLYIKQEEFSPSONSQPGL
            LVRAQHEGVRTAGAVOEGSIIRGTTPASKIVETISLRSSTTQGTPLPAQAGIPTE
            ALVKGPVSNPIEESPEKVRSEAKSHVIEGKSGHTLSYDLSLCRALPRGPHSD
            LKERTVLSIMQGTPTAESEFDGLKPKQIKRESPIRAPEGAITKGPYDGIIT
            IKEMGSIHEIPRODILTQESRKTPEVQSTPIEGSISQGTPIKFDNNSSQSAIKH
            NVKSLITGSPKLMLEIVPENIKVVERGTDVKAGEPVRAHRTSVVSSSGSVLRS
            TLHEAPKALSPGLYDSDSARTPVSYQNTISRGSPMMNRTSDVSSKASHERKSTL
            TTPQRESIPAKSPVGVDPVSHSPDPHRSAGEVVRSHLPTHLDPAMPFRHALD
            PAAAILQRQLSPTEGYSQQLYAMENTROTILNDYITSOQMVNLAPDVTGRLSP
            REQPLGLPVAIRGIIIDLNTMPTLIVPHAGGSTPTPMDRITVITGTVTFPPPYNA
            ASLSPGHPTHLAASAERERERERERERERERERERERERERERERERERERERER
            QPGRGSHGVYRSPSPVSTQETILQORPSPQGTNGTSTVITLPTAQLRMLPSG
            GPSISQGLPASRVNTAADAALVDAAPQMDVSKTKESKHEAARLEENLRSGAA
            VSEQQOLEOKNLVEKRSVCQCTSSALSGKAQPHASVYSEAGKDGPPKPSRYEE
            ELTRGKTTITANFIDVILITQIASDKDAREGSSDSSSSLSHRYETASDAIEV
            ISPASSFAPPQKFPQAIQFDMVKANAENESTRQYEGPLHHYRSQSESPPQPPLP
```

```
gene
CDS
```

595	AGGCCTGAGCCA	CAGTCAGACGACGACTCAGTGCACCTTGCGAGTCCGATAGGGGTGTG	654
928	GluaAlaGluGlyGlyAspLysAsnArgLeu	-----LeuSerProArgProSerLeuLeu	945
655	GATGGAGAG--	CCAGAGAGCAGAGGTGTTCCCATGATGATCAAGCCTTCATTGTTA	711
946	ThrProThrGlyAspProArgAlaAsnAlaSerProGlnLys	-----ProLeuAspLeu	963
712	ACTCCTCCTGGATCTATCCTAATC--	TCATCCCTATTAAACCAACCCATGGATCTG	768
964	LysGlnLeuLysGlnArgAlaAlaAlaProProle	-----	976
769	CCACAGCTTCAGATCGAGCTGCTGTATTCCACCAATGGTTCTTGCACCTCCAGTAAT	828	
977	-----	GlnValThrLysVal	981
829	ATACCAATTGGAAAGCCCGTAAGTGGCTATGCTCTTTACCAACGGCACATTAAGGCCATG	888	
982	HieGluProProArgGluAspAlaAlaProThrLysProAlaProProPro	1001	
889	CATGAGTCAGCACTCCTGGAG-----	GAG	912
1002	ProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArgGly	1021	
913	CACGCGCAGAGGCAAGACAGTAGACTTGGAAATGCAGAAAGCTCTCAAGACCCATGCACG	972	
1022	LysSerArgSerProAlaProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLys	1041	
973	ACTTCTAAGAGTCCA-----	ACAGGGAG-	996
1042	LeuProGlyAspProCysThrThrSerGlyLeuPheProValProArgGlu	1061	
997	-----	TGG-----GAA	1002
1062	ValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaProProGly	1081	
1003	GTCTCTCAACCTCTCCGCAT-----	CAAGTGATAACTAACCTCTCGTAAGGG	1050
1082	HisProLeuGlyLeuHisAspThrAlaArgProValLeuArgProProThr	1101	
1051	GTTCGGGTTCGA-----	ACACACGACCAACAGGCCACACCTCCCTC	1095
1102	IleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGln	1121	
1096	ATCCATCATCTATAAACACACAGTGGCATCAGAAAA--	CCATCCTTTATA-----ATG	1146
1122	IleGlyAlaIleSerGlnGly-----	MetSerValGlnLeuHisValPro	1136
1147	GGAGGGTCTATCTCAACAGGGAACCTCTGGCACTTACTTGTCTCTCATTAATCAGGCTTAT	1206	
1137	TyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMet	1156	
1207	CCACAGAGACCCCTAAGCCCTCCGCGGGGTCTATCTCTCGGATTTGCCCGCAGCAG	1266	
1157	AspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGly	1176	
1267	GAGTCTACCAAGCAGCTCTTTGACCTACATCAAGCAGGAAGAATTTTCTCGAGAACG	1326	
1177	GlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArg	1196	
1327	CAAAACTCAACACTGAGGGTTTATTGGTCT--	AGAGCGCAGCATGAAGTGTGGTCTAGA	1383
1197	GlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArg	1216	
1384	GGCACTGTCA--	GGGGCTGTCCAAAGAAAGAGATTAACTCGGGGAACTCCACCGACAAA	1440
1217	ValProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAla	1235	
1441	ATCTCAGTGGAGACCATTCATCGTTCGGGGCTCTATTACCAGGGGACCCAGCTCTG	1500	
1236	-----	AspValLeuTyrLysGlyThrIleThrArgIle	1247
1501	CCCCAGGCTGGAATACCAACAGAGGGTTTGGTGAAGGACCTGTCTCCAGGATGCTATT	1560	

QY 1248 GlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHis 1267
Db 1561 GAAGAAGAGCATCTCTGAGAGGTC-----AGAGAGAGAGCTCCATCCAGGGCCAT 1611
QY 1268 ValIleArgGlyLysGlyHisValLeuSerTyrGluGlyMetSerValThr 1287
Db 1612 GTTATCTATGAGGCAAAAGTGACATATCTATCATAT----- 1650
QY 1288 GlnCysSerLysGluAspGlyArgSerSerGlyProProHisGluThrAlaAlaPro 1307
Db 1650 ----- 1650
QY 1308 LysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIle 1327
Db 1650 ----- 1650
QY 1328 GluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHis---HisLeuLys 1346
Db 1651 GATAGTCTGATATGCCGAGCATATCCC-----AGGGGAGCCCTCATTTCTGACCTCAAA 1704
QY 1347 GluGlnHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAla 1366
Db 1705 GAGAGGACTGTGTCTGTCTGTTCCATAATGCAGGCGCACCAAGAGCCACAGCGGAAGC 1764
QY 1367 GlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProPro 1386
Db 1765 TTTGAGGACCGCCTT---AAATACCCCAACACATAAAAGGAGAGC----- 1809
QY 1387 ProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLys 1406
Db 1810 ---CCTCCCATCCGAGCATTTGAGGTGCCATT-----ACCATA 1845
QY 1407 LeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHis 1426
Db 1846 GGAATAACCA---TACGATGT---ATCACCAACATCAAAAGAAATGGCGCTCCATCCAT 1899
QY 1427 GluIleProArgGluGluLeu-----ArgHisThrProGluLeuProLeu 1441
Db 1900 GAAATCCACGGAAGATATTTCACTCAGAAAGCCGGAATACTCCAGAAAGTGTCCAG 1959
QY 1442 AlaProArgProLeuLysGlySerIleThrGlnGlyThrProLeuLysTyrAspThr 1461
Db 1960 AGCAGAGGCCAATAATTGAGGTTCATTTCACGGGCACACCAATAAAATTTGACAAC 2019
QY 1462 GlyAlaSerThrThrGlySerLysLysHisAspValArgSerIleGlySerProGly 1481
Db 2020 AAC---TCAGGTCAATCAGCTATCAACACAAATGTGAAGTCTTAATCAGAGGCGCTAGC 2076
QY 1482 ArgThrPheProProValHisProLeuAspValMetAlaAsp---AlaArgAlaLeuGlu 1500
Db 2077 AAACATA-----CCCCGTGAATGCTGGAATGTTCACGAGAACATAAAGTAGTAGAA 2130
QY 1501 ArgAlaCysTyrGlu-----GluSerLeuLysSerArgProGlyThrAla 1515
Db 2131 CGGGGAAATATGAGGATGTGAAGAGCGCGCAGTCAGTGGAGCCGCGCACAGTCAGTG 2190
QY 1516 SerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLys 1535
Db 2191 GTGAGCTCTGGCCCTCGTTCTCAGGTCTACA-----CTTCAGAAAGTCCCAAA 2241
QY 1536 ProArgGlnSerProLeuThrTyrGluAspHisGlyAla-----ProPhe 1550
Db 2242 GCACAGTACCGCCGAGCTCTATGATGACAGAGCGCTCGCAGGACTCTCTGTGAGCTAC 2301
QY 1551 AlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGln 1570
Db 2302 CAGAACCATATCCAGAGGCTCCCCCATGATGAACAGA----- 2340
QY 1571 GluGlySerLeuSerSerSerLys---AlaSerGlnAspArgLysLeuThrSerThrPro 1589
Db 2341 ACTTCTGATGTTCTTCCAGCAAGTCTGCCAGTCAATGAAAGGAATAACATCTGACCCCA 2400

QY 1590 ArgGlu-----IleAlaLysSerProHisSerThrValProGluHisHisPro 1605
Db 2401 ACCCAAGAGAAAGTATACCAAGCAAGTCTCCAGTGCAGCGGGTGGATCCCGCTCGTAGT 2460
QY 1606 HisProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArg 1625
Db 2461 CAC-----AGCCCATTTGATCTCTCATCAGAGTAGCGCTGCAGGAGAGGTTTATCGG 2514
QY 1626 SerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAsp--- 1644
Db 2515 AGCCACCTTACCACCGACTTGGATCCA---GCTATGCCCTTTTCACAGGCTTGGATCT 2571
QY 1645 AlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeu 1664
Db 2572 GCAGCTGTCTTACTCTGTTCACAGACAGCTTTTACCAACCCAGGATACCAAGTCAG 2631
QY 1665 TyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsn---ArgGln 1683
Db 2632 TACCAGCTCTAT-----GCAATGGAGAAATACAAAGCCAG 2664
QY 1684 ThrIleLeuAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMet 1703
Db 2665 ACAATCTCAAGATTTACATTACCTCACAGCAGATGCAGTGAATCTG----- 2712
QY 1704 AlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsn 1723
Db 2713 -----CGCCCTGATGTACACAGGAGCTGTCCACAGAGAGCAGCACTGGGGCTCCCT 2766
QY 1724 TyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeu 1743
Db 2767 TACCCAGCT---ACAAGAGGAATCATTTGACCTGACCAATATGCCTCCA---ACAATCTTA 2820
QY 1744 ValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThr 1763
Db 2821 GTGCCTCATGAGGGGAACGAGCACCCCTCCATGACAGCATCAGTATATCTCTGTT 2880
QY 1764 AlaProGlnProPheSerSerArg---HisSerSerSerProLeuSerProGlyGlyPro 1782
Db 2881 ACACAGGTTACTTTCCCTCCAGGCCATATACGCTGTCTCTCTGCTCCAGACACCCA 2940
QY 1783 ThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArg 1802
Db 2941 ACACACCTT-----GCACGAGCTGCAAGTGTCTGAGAGGAACGAGAGCGGAAGG 2991
QY 1803 GluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThrValGlu 1822
Db 2992 GAGAGGAGCGCAACGTCGAGCGGAGCGTGCAGCTGAGCTGAACGTCGAAACGCGAAGATCGCT 3051
QY 1823 HisAlaPro-----IleTrpArgProGlyThrGluGlnSerSerGlySerSerGly 1839
Db 3052 GCTGCTCCCGCTGACCTCTACCTTACCTACGACAGGTTTCAGAACAG----- 3093
QY 1840 SerSerGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGln 1859
Db 3094 -----CCAGCGCGTCTGGGAGCCAGCCAGGATATGTTCCGCTCC 3129
QY 1860 HisSerProIleSerProArgThrGlnAspAla---LeuGlnGlnArgProSerValLeu 1878
Db 3130 CCTTCCCT---TCAGTAGAAGTCTCAGGAGACCATCTCTCAACAGAGACCCAGTGTTC 3186
QY 1879 HisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeu 1898
Db 3187 CAGGGCACCAATGGAACACCATGTAAATCACACCTTTGGACCCCACTGCTCAGCTACGCATC 3246
QY 1899 ArgSerThrSerThrSerSerProVal-----ArgProAlaAlaThrPhePro 1914
Db 3247 ATGCCACTGCTTCTGGGGGCCCTTCCATAAGTCAAGGCCCTGCAGCCTCCCGTTACAC 3306
QY 1915 ProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGlu 1934
Db 3307 ACTGCTCGGATGCC---CTGGCTGCTCTGTGATGCTGCAGCTTCTGCACCCAGATG 3363
QY 1935 ProValLeuLeuProLysGluAlaProArg----- 1944

Db	3364	GATGTTTCCAAAACAAAGAGAGTAAGCATGAAGCTGCCAGGTTAGAAGAAAAATTTCGAGA	3423
Qy	1945	-----ValAlaIArgProGluArgAlaAspThrGlyHisAlaPheLeuAla	1961
Db	3424	AGCAGGTGCAGCAGCAGTTAGTAACAGCAGCAGCTAGACAGAAAAACCTGGAGGTGGAG	3483
Qy	1962	LysProProAlaArserGlyLeuGluProAlaSerSerProSerLysGlySerGluPro	1981
Db	3484	AAGAGACTGTGTTCAGTGTTTTGCACCTCTTCAGCCCTTCCAAGTGCAGCAAGGCCACGCT	3543
Qy	1982	ArgProLeuVal-----ProProValSerGly	1990
Db	3544	CATCGCTCAGTAGTGTTATCTGAGGCTGGGAAGATAAAGGGCTCTCTCCAAATCCAGA	3603
Qy	1991	HisAlaThrIleAlaArgThrProAlaIAsn-----LeuAlaProHisHisAlaSer	2008
Db	3604	TATCAGGAAGAGCTAAGGACCCGAGGAGAGTACCATTACTGCAGTAACCTCATCATGAC	3663
Qy	2009	ProAspProAlaIProProAlaSerAlaSerAspProHisArgGluLysThrGlnSer	2028
Db	3664	GTGATCATCACCCGGCAAATTGCCTCGGACAAGATCGAGGGAAACGTGCTCTCTCAAGT	3723
Qy	2029	LysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyHisGlySerSerTy	2048
Db	3724	TCAGACTCTTCT-----AGTAGCTTGCTTCTCTCACAGGTATGAAACG	3765
Qy	2049	SerProGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLys	2068
Db	3766	GCTAGTGATGCATTGAGGTGATTAAGTCCCGCCAGCTCACCTGCACACCACCCAGGAAAAG	3825
Qy	2069	GlyLeuProLysHisLeuGluLeuLeuAspLysSerHisLeuGluGlyGluLeuArgPro	2088
Db	3826	CCACAGSCCTATCAGCCAGACATGGTTAAGCGAAATCAAGCAGAAAATGAGTCCACTCGA	3885
Qy	2089	LysGlnProGlyProValLysLeuGlyGlyGluAlaHisLeuProHisLeuArg---	2107
Db	3886	CAGTAAAGGTCCA-----CTGCATCATATTATCGGTCC	3918
Qy	2108	-----ProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThr---	2122
Db	3919	CAGCAGGAATCACCATTCCACAGCAACAGCCACCATCGCCCCCATCTCCCAGTCAGAG	3978
Qy	2123	---AlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGlu	2141
Db	3979	GGAATGGACAGGTGCCAGGACCCATCGACTGATCACACTTGTGTGACACACATCTGTCAA	4038
Qy	2142	ValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAla	2161
Db	4039	ATTATCACAAAGATTTTGTAGAAAT-----CAAGTTCCCTCGCAGCCTTCTACTTCT	4092
Qy	2162	ProLeuTyrSerPheProGlyAla-----SerCysProValLeuAspLeuArgArgPro	2179
Db	4093	ACATTCCAAATCTCACCATCTGCTTTTGTATCCACACCTGTA-----AGAATCAA	4143
Qy	2180	ProSerAspLeuTyrLeuProProProAsp-----HisGlyAlaProAla	2194
Db	4144	ACCTCAGCCGCTACACCCAGAATCACAGTCTCAGACTCTCTGTGATCCAGACCAGGT	4203
Qy	2195	---ArgGlySerProHis-----SerGluGlyLysArgSerProGlu	2208
Db	4204	CCTAGAGTCTCTCCAGAAAAATCTTGTGGATAAATCCCGGGAAGCAGGCGCTGGAAAATCT	4263
Qy	2209	ProAsnLysThrSerValLeuGlyGlyGluAspGlyIleGluProValSerProPro	2228
Db	4264	CCAGAGAGGAGTCATATC-----CCATCAGAGCCCTATCAGGCCATCTCCCCACCC	4314
Qy	2229	GluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAsp	2248
Db	4315	CAAGGCCCTCTGTG-----CATGAAACAGCAGCAGCATGTGTCTTGTCTACAGAGG	4368
Qy	2249	GlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnPro	2268

Db	4369	GGAGTGACCTGCTGTAGCAAGAGAGTGATTCTCGATCACCGAAGAGTATAAGCTACTTG	4426
Qy	2269	ProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGln	2288
Db	4429	CCTTCATCTTCCACAGCTT--GAAGGACATACCCATGGTTAATCAAGAACAACAG	4485
Qy	2289	GluLeuAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluThrYrAsnLysSerGln	2308
Db	4486	GAATTTTTCCTAAGTTGAATCTCTCTGTTGGAGGTGACTCTGATATGGCAGCTGCTCAG	4545
Qy	2309	ProGlyThrGluThrPheAsnMetProAlaIleThrGlyThrGlyLeuMetThrYrArg	2328
Db	4546	CCAGGAACAGAGATCTTCAATCTGCCAGAGTTACCAATCATGAGTGGAGTCAAGA	4605
Qy	2329	SerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleLeuArgLys	2348
Db	4606	AGCATTTCTTTGCTGATCCGGCAGT--AACCTTGGTCTAGNAGACATCATCAGAAG	4662
Qy	2349	AlaLeuMetGlyLysTyrAspGlnTrpGluGlu-----SerProPro	2362
Db	4663	GCTCTCATGGGAAGTTTGTATGATAAAGTTGAAGATCATGGTGTGTCATGTGCCACTCT	4722
Qy	2363	LeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetPro	2382
Db	4723	GTG-----GSCATATTAGCTGGTGGTGGCCAGCACCTCAGTG-----	4758
Qy	2383	IleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGly---Gly	2401
Db	4759	GTGACGAGCAGCGGACCGAGAGATGAAGGGAGCCATCCTCATGCGAGGATGTC	4818
Qy	2402	LysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeu	2421
Db	4819	AAACCAAGCTGATCAACAAATCAACAGCAGGAAAGTCTAAATCTCTTCTCTGGGCAA	4878
Qy	2422	Ala-----SerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCys	2439
Db	4879	AGCTATTAGAACTGAAGAGCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	4938
Qy	2440	AsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySer	2459
Db	4939	CACAGGCAGACCA-----GGATGGCATGGAGATCGGCCCTCTTCAACAGGTCTCT	4992
Qy	2460	ThrProPheProTrpAsnProLeuLeuMetArgLeuGlnAlaGlyValMetAlaSerPro	2479
Db	4993	ACTCAGTTCCCTTACACCTCTGACCATACGATG-----CTCAGCAGTACA	5040
Qy	2480	ProProProGlyLeuProAlaGlySerGlyProLeu-----AlaGlyProHis-----	2495
Db	5041	CCACCTTACAGATCGCATCGCCCTCTGCCATCATCCCAAGCAGCTCCACATCAACAG	5100
Qy	2496	---HisAlaTrpAspGluGluProLysProLeuLysCysSerGlnTrpGluThrLeuSer	2514
Db	5101	AACGGCATCTGGAGAGGAGGAGCTGCCCCCTCTCTCTCAGCGCAGTATGACACTGCTCT	5160
Qy	2515	AspSerGlu	2517
Db	5161	GATAGTGAC	5169
RESULT	24		
LOCUS	AB019524	6541 bp	linear
DEFINITION	Homo sapiens mRNA for nuclear receptor co-repressor, partial cds.		PRI 19-AUG-1999
ACCESSION	AB019524		
VERSION	AB019524.1	GI:4519939	
KEYWORDS	nuclear receptor co-repressor.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (sites)		
	Nagaya,T., Chen,K.S., Fujieda,M., Ohmori,S., Richer,J.K.,		
	Horwitz,K.B., Lupeki,J.R. and Seo.H.		
	Localization of the human nuclear receptor corepressor (hN-Cor)		

Qy 1385 ProProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyPro 1404
Db 1048 -----CCTCCATACGAGCAATTTCAAGGTGCCATT----- 1077
Qy 1405 LeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSer 1424
Db 1078 ACCAAGGAAACCA---TATGATGCG---ATCCACCAATCAAGAAATGGGGCGTTCC 1131
Qy 1425 IleHisGluIleProArgGluGluLeu-----ArgHisThrProGluLeu 1439
Db 1132 ATTCATGAGATTCACAGCGAAGATATTTAACTCAGGAAGTCCGAAATCTCCAGAAGTG 1191
Qy 1440 ProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyr 1459
Db 1192 GTCCAGAGCACACGGCGCAATATGAGGTTCATTTCCAGGGCACACCAATAAAGTTT 1251
Qy 1460 AspThrGlyAlaSerThrGlySerLysHisAspValArgSerLeuIleGlySer 1479
Db 1252 GACAACAAC---TCAGTCAATCTGCCATCAACACCAATGTCAATCTCTTAATCACCGGG 1308
Qy 1480 ProGlyArgThrPheProProValHisProLeuAspValMetAlaAsp---AlaArgAla 1498
Db 1309 CCTAGCAAACTATCCCGTGAATGCTCCGCTGGAATTTGTCCAGAGAACATAAAGTG 1368
Qy 1499 LeuGluArgAlaCysTyrGlu-----GluSerLeuLysSerArgProGly 1513
Db 1369 GTAGAACGGGMAAATATGAGGATGTGAAGACGGCGAGACCGTGCCTCCCGGCACACG 1428
Qy 1514 ThrAlaSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeu 1533
Db 1429 TCAGTGTGTAAGCTGTGCCCTCCGTTCTTAGGTCCACA-----CTGCATGAAGCT 1479
Qy 1534 GlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAla----- 1548
Db 1480 CCNAGCACACTGACCCCTCGGATTTATGATGACACCACTGCACGGAGACCCCTGTG 1539
Qy 1549 ProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrProArg 1568
Db 1540 AGTTATCAAAACACCATGTCCAGAGGCTCACCCATG-----ATGAACAGA 1584
Qy 1569 LeuGlnGluGlySerLeuSerSerLysAla---SerGlnAspArgLysLeuThrSer 1587
Db 1585 ACTTCTGATGTACAATTTCTTAAACAAGTCTACCAATCATGAAGGNAATCGACACTG 1644
Qy 1588 ThrProArgGlu-----IleAlaLysSerProHisSerThrValProGluHis 1603
Db 1645 ACCCTACCAGAGGGAAGTATCCAGCGAGTCTCCAGTGCCTGGGGTGGACCTGTG 1704
Qy 1604 HisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeu 1623
Db 1705 GTGAGCCAC-----AGTCCGTTTGATCCCATCCATCACAGAGCGACACTGCAGGCGAGGTT 1758
Qy 1624 TyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeu 1643
Db 1759 TATCGGAGCACCTGCGCCACCACTTGATGCCA---GCCATGCCCTTTTCAGGGCTTTG 1815
Qy 1644 Asp---AlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrPro 1662
Db 1816 GATCCTGCAGGGCTGTCTACCTGTTTCAGAGACAGCTTTCACCACTCCAGGTACCCA 1875
Qy 1663 HisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsn--- 1681
Db 1876 AGTCAGTATCAGCTTTAC-----GCAATGGAGAACACA 1908
Qy 1682 ArgGlnThrIleLeuAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThr 1701
Db 1909 AGACAGACAATCTTAAATGATTACATTACCTCAACACAGATGCAAGTGAACCTTG--- 1962
Qy 1702 AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAla 1721
Db 1963 -----CCTCAGATGTGGCCAGGAGGACTCTCCCAAGAGAGACCCACTGGT 2010
Qy 1722 LeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuPro 1741

Db 2011 CTCCCATACCCAGCA---ACGAGAGGAATCATTTGACCTGACCAATATGCTCTCCA---ACA 2064
Qy 1742 ValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeu 1761
Db 2065 ATTTAGTGTCTCATCCAGGGGAACAAGCACTCTCCCATGGACAGAACTTATAT 2124
Qy 1762 ProThrAlaProGlnProPheSerSerArg---HisSerSerProLeuSerProGly 1780
Db 2125 CCTGGTACACAGATTACTTTCCCTCCAGCGCGGTACAACTCTGCTTCCATGTCTCCAGGA 2184
Qy 1781 GlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArg 1800
Db 2185 CACCCAAACACACTT-----GCAGCTGCTGCAAGTGTGAGAGGAACCGGAACGG 2235
Qy 1801 AspArgGluArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThr 1820
Db 2236 GAGCGGGAAGAGAGCGGAGCGGAACGG-----ATTGCTGCAGCTTCTCTCCGAC 2286
Qy 1821 ValGluHisAlaProIleThrArgProGlyThrGluGlnSerSerGlySerGlySer 1840
Db 2287 CTC-----TACCTGCGGCCAGGCTCAGAACAG----- 2313
Qy 1841 SerGlyGlyGlyGlySerSerArgProAlaSerHisSerHisAlaHisGlnHis 1860
Db 2314 -----CCTGCGCGACCTGGCAGTCTGATATGTTGCTCCCT 2352
Qy 1861 SerProIleSerProArgThrGlnAspAla---LeuGlnGlnArgProSerValLeuHis 1879
Db 2353 TCCCT---TCAGTAGAACTCAGGAGACCATGTTGCAACAGAGACCCAGTGTTCCTCAA 2409
Qy 1880 AsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArg 1899
Db 2410 GGAACCAATGGAACCATGTAATCACACCTTGGATCCAACTGCTCAGTACGATCATG 2469
Qy 1900 SerThrSerThrSerSerProVal-----ArgProAlaAlaThrPheProPro 1915
Db 2470 CCAGTGCCTGCTGGGGCCCTTCAATAAGCAAGCGCTGCCAGCTCCCGTTACAACACT 2529
Qy 1916 AlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluPro 1935
Db 2530 GCTCGGATGCC---CTGGCTGCTTGTGGATGCTGCAGCTTCTGCACCCAGATGAT 2586
Qy 1936 ValLeuLeuProLys-----GluAlaProArgValAlaArgProGluArgPro 1951
Db 2587 GTGTCCAAACAAAGAGAGATGAGCATGAGCTGCCAGGTAGAGAAGAAATTTGAGAAGC 2646
Qy 1952 ArgAlaAsp----- 1954
Db 2647 AGTCTACGACAGCTTAGTGAACAGCAGCAGCTAGAGCAGAAAAACCTCGAGGTGGAGAG 2706
Qy 1955 -----ThrGlyHisAlaPheLeuAlaLysProProAlaArgSer 1967
Db 2707 AGATCTGTTGAGTGTATACACTTCTTCCAGCTTTTCCAAAGTGGCAAGCCC----- 2757
Qy 1968 GlyLeuGluProAlaSerSer-----ProSerLysGlySerGluProArgProLeu 1984
Db 2758 -----CAGCTCATCTTCTCAGTAGTTTATTCTGAGCTGGGAAAGATAAAGGSCCT--- 2808
Qy 1985 ValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsn-----Leu 2002
Db 2809 ---CCTCCAAATCCAGATATGAGGAAGAGCTAAGGACCCAGAGGGAAGACTTACCTACT 2865
Qy 2003 AlaProHisHisAlaSerProAspProProAlaProProAlaSerAspProHis 2022
Db 2866 GCAGTAACCTTATAGACGTGATCATCCCGGCAATTTGCTCGGACAGGATGCGAGG 2925
Qy 2023 ArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluArgSerLeuGly 2042
Db 2926 GAAGTGGCTCTCAAGTTCAGACTCTTCT-----AGTAGCTTATCT 2967
Qy 2043 TyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSerSerPro 2062

2968 TCTCACAGGTATGAAACACCTAGCGATGCTATTAGGTGATTAAGTCTCTGCCAGCTCACT 3027 Db
2063 SerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeu 2082 Qy
3028 GCCCACCAGAGAAATCGCAGACTATCAGCAGAGGTTGTTAAGCAAAATCAAGCG 3087 Db
2083 GluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaHis 2102 Qy
3088 GAAATGATCTCTACCAGACAATATGAAGACCA----- 3120 Db
2103 LeuProHisLeuArgProLeuProGluSerProLysSerSerProLeuLeuGlnThr 2122 Qy
3121 TTATCATCACTATCGACA-----CAGCAGAAATCCATCTCTCCCAACAACAGCTG 3171 Db
2123 AlaProGly-----ValLysGlyHisGlnArgValValThr 2134 Qy
3172 CCCCCTTCTTCACAGGAGGGAATGGGCAAGTCCCGAGGACCATCGGCTGATCACA 3231 Db
2135 LeuAlaGlnHisLeuSerGluValIleThrGlnAspTyrThrArgHisHis----- 2151 Qy
3232 CTTGCTGATCACAATCTCTCAAAATATCACAAGATTTTGCTAGAAATCAAGTTTCTCG 3291 Db
2152 -----ProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAla 2169 Qy
3292 CAGACTCCCGAGCAG-----CCTCTACTTCTACATTCAGAACTCACCTTCTGCT 3342 Db
2170 -----SerCysProValLeuAspLeuArgProProSerAspLeuTyrLeuProPro 2187 Qy
3343 TTGGTATCTACACTGTG-----AGGACTAAACATCAAAACCGTTACAGCCAGAA 3393 Db
2188 ProAsp-----HisGlyAlaPro-----AlaArgGlySerProHis----- 2199 Qy
3394 TCCAGGCTCAGTCTGTCCATCAACAGAGAGTTCAGGGTCTCTCCAGAAATCTT 3453 Db
2200 -----SerGluGlyGlyLeuArgSerProGluProAsnLysThrSerValLeuGly 2216 Qy
3454 GTGACAATTCAGGGGAGTAGGCTCGAAATCCCGAGAGGAGTCAAGTCTC----- 3507 Db
2217 GlyGlyGluAspGlyIleGluProValSerProProGluMetThrGluProGlyHis 2236 Qy
3508 ---TCTCGAGCCCTACAGGCCATCTCCCGACCCAG-----GTTCCGGTGTGTCAT 3558 Db
2237 SerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArg 2256 Qy
3559 GAGAAACAGGACAGCTTCTGCTCTGCTCAGAGGGCGCAGAGCTCGCAGACGAGG 3618 Db
2257 MetGlySerLysSerProGlyAsnThrSerGlnProProAlaPheSerLysLeuThr 2276 Qy
3619 AATGATGCCCGCTCACCAGGAGTATAAGTACTTGCCTTCATCTTCCACCAAGCTT--- 3675 Db
2277 GluSerAsnSerAlaMetValLysSerLysGlnGluIleAsnLysLysLeuAsnThr 2296 Qy
3676 GAAATAATACATCCCATGTTAAATCAAGAAGCAGGAGATTTTCGTAAGTTGAATCC 3735 Db
2297 HisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMet 2316 Qy
3736 TCTGGTGGAGGTGATCTGATATGCGAGTCTGCTCAGCCAGGACATGAGATCTTTAATCTG 3795 Db
2317 ProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAla 2336 Qy
3796 CCAGCAGTTACTACGTCAGGCTCAGTTAGCTCTAGAGGCCATCTTTTGTGATCCTGCC 3855 Db
2337 SerThrAsnMetGlyLeuGluAlaIleArgLysAlaLeuMetGlyLysTyrAspGln 2356 Qy
3856 AGT---AATCTTGGGCTGGAAGACATTATCAGAGAGGCTCTCATGGAGCTTTGATGAC 3912 Db
2357 TrpGluGlu-----SerProProLeuSerAlaAsnAlaPheAsnPro 2370 Qy
3913 AAGATTGAGGATCATGGATTGTCTATGTCCTCCAGCTTATGGAGTA-----GTGCT 3963 Db
2371 LeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSer 2390 Qy
3964 GGTACTGCCAACACCTCAGTT-----GTGACCAGGTGGTGACAGACAGAA 4008 Db

2391 AspHisThrLeuThrSerProGlyGlyGlyGly-----LysAlaLysValSerGlyArg 2408 Qy
4009 GAGGAAGGGAGCCCATCACCTCATTACAGGAGGAGTTTGCAAAAGCTGATCAGCAAG 4068 Db
2409 ProSerSerArgLysAlaLysSerProAlaPro-----GlyLeuAlaSerGlyAspArg 2426 Qy
4069 TCAACAGCAGCAAAATCTAGTCTCTTATACCTGGCAAGGCTACTTTAGGAACGGAACGG 4128 Db
2427 ProProSerValSerSerValHisSerGluClyAspCysAsnArgArgThrProLeuThr 2446 Qy
4129 CCCTCTTCAGTCTCTCTGTATCATTCAGAGGGGATTCACATAGCAGACGCCA----- 4182 Db
2447 AsnArgValTyrGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnPro 2466 Qy
4183 GGTGGGGCTGGGAGACAGAGCCCTCTTCAACAGGCTCACTCAGTTCTCTTATAGCCT 4242 Db
2467 LeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAla 2486 Qy
4243 CTGACTATGCGATG-----CTCAGCAGTACTCCACCAACACCGATTGCGATGCT 4293 Db
2487 GlySerGly---ProLeuAlaGlyProHis-----HisAlaTrpAspGluGluPro 2502 Qy
4294 CCCTCTGGGGTGAACCAAGCAGCTCTCCCAACAGACAGATCTGGGAGGAGGCT 4353 Db
2503 LysProLeuLeuCysSerGlnTyrGluThrLeuSerAspSerGlu 2517 Qy
4354 GCCCACTGCTCTCAGCACAGTACGAGACCTGTGCGATAGTGAT 4398 Db

RESULT 25
MUSRIPI3G
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

MUSRIPI3G 4285 bp mRNA linear ROD 10-DEC-2001
Mus musculus RIP-13 (RIP13) mRNA, complete cds.
L78294
L78294.1 GI:17467267
RIP13 gene; corepressor; n-cor gene; nuclear hormone receptor.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Seol,W., Mahon,M.J., Lee,Y.K. and Moore,D.D.
Two receptor interacting domains in the nuclear hormone receptor
corepressor RIP13/N-Cor
Mol. Endocrinol. 10 (12), 1646-1655 (1996)
97120602
8961273
RIP13 is identical to the N-Cor, except that it has 10 unique amino
acids at its N-terminus instead of N-Cor's 1017 amino acids of its
N-terminus and contains a 48 amino acids deletion (N-Cor No.
1235-1282). In addition, RIP13 has one more alanine (N-Cor 1589)
and a serine deletion (2145) and a amino acid change from A to P.
These minor differences may be due to polymorphism. RIP13 sequence
has been reconstructed using three clones whose sequences are
overlapped partially. There is a deletion form of RIP13 containing
a deletion of 119 amino acids (RIP13delta).

Location/Qualifiers
1. .4285
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/tissue_type="liver"
1. .4285
/gene="RIP13"
1. .40
/note="putative"
40. .4239
/gene="RIP13"
/note="putative"
/codon_start=1
/product="RIP-13"
/protein_id="AAL40135.1"

FEATURES
source
gene
5' UTR
CDS

Db 1294 CACGAAGCTCCCAAGACGACGAGCTGAGCCGGGACTCTATGATGACAGCAGCGCTCGCAGG 1353
Qy 1549 -----ProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluPro 1565
Db 1354 ACTCTGTGTGCTACCAAGACCATATCCAGAGGCTCCCCATGATGAACAGA----- 1407
Qy 1566 ThrProArgLeuGlnGluGlySerLeuSerSerSerLys---AlaSerGlnAspArgLys 1584
Db 1408 -----ACTCTGTGATGTTCTTCCAGCAAGTCTCCAGTCCAGTATGAAGAGAA 1452
Qy 1585 LeuThrSerThrProArgGlu-----IleAlaLysSerProHisSerThrVal 1600
Db 1453 TCACCTGTGACCCCAACCAAGAGAGATATACCAAGCAAGTCTCCAGTCCCGGGTG 1512
Qy 1601 ProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerGly 1620
Db 1513 GATCCCGCTGTGATCAC-----AGCCCATTTGATCCTCATCACAGGAGTAGCGTGCA 1566
Qy 1621 ValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGly 1640
Db 1567 GGAGAGGTATTATCGAGGCCACCTACCCACCGACTTGGATCCA---GCTATGCCCTTTTAC 1623
Qy 1641 IleProLeuAsp---AlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnPro 1659
Db 1624 AGGCTTTGGTCTGCTGAGTGTCTTACCTGTACAGAGACAGCTTTTCAACACCCCA 1683
Qy 1660 ThrTyrProHisLeuTyrProTyrTyrLeuIleArgGlyTyrProAspThrAlaAlaLeu 1679
Db 1684 GGATACCCCAAGTCAGTACCAGCTCTAT-----GCAATG 1716
Qy 1680 GluAsn---ArgGlnThrIleLeuAsnAspTyrIleThrSerGlnGlnMetHisHisAsn 1698
Db 1717 GAGAAATACAGGCAGACAACTCTACGATTTACCTTACCTACAGAGACAGCTTTTCAACACCCCA 1776
Qy 1699 ThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSer 1718
Db 1777 CTG-----CGCCCTGATGTCACACAGGAGGAGCTGTCACACAGAGCAG 1818
Qy 1719 SerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleLeuAspLeuSerGlnValPro 1738
Db 1819 CCACTGGGGCTCCCTTACCCAGCT---ACAAGAGGAATATTGACCTGACCAATATGCT 1875
Qy 1739 HisLeuProValLeuValProThrProGlyThrProAlaThrAlaMetAspArgLeu 1758
Db 1876 CCA---ACATCTTAGTGCCTCATGCGGGGGAACGAGACCCCTCCCATGGAGCATC 1932
Qy 1759 AlaTyrLeuProThrAlaProGlnProPheSerSerArg---HisSerSerSerProLeu 1777
Db 1933 ACGTATATTCCTGTGTACACAGGTTACTTTCCTCCCGGCCATATAACGCTGCTTCTCTG 1992
Qy 1778 SerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGlu 1797
Db 1993 TCTCCAGGACACCAACACACTT-----GCAGCAGCTGCAAGTGTGAGAGGGAA 2043
Qy 1798 ArgAspArgAspArgGluArgAspArgGluArgGluLysSerIleLeuThrSer 1817
Db 2044 CGAGAGCGGGAAGAGGAGAGCGGAGAGCTGAGCGGAGCTGAGCGGTGAACGTGAA 2103
Qy 1818 ThrThrThrValGluHisAlaPro-----lleTrpArgProGlyThrGluGlnSer 1834
Db 2104 CGCGAAAGGATCGTGTGCTCCCGCTGACCTCTACCTACGACGAGTTCAGAACAG--- 2160
Qy 1835 SerGlySerSerGlySerSerGlyGlyGlyGlySerSerArgProAlaSerHis 1854
Db 2161 -----CCAGGCGGTCTTGGCAGGCCAC 2181
Qy 1855 SerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAla---LeuGlnGln 1873
Db 2182 GGATATGTTCCGCTCCCTTCCCT---TCAGTAAGAACTCAGGAGACCATCTCTCAACAG 2238
Qy 1874 ArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSer 1893
Db 2239 AGACCCAGTGTTTTCCAGGGCCACCAATGGAACCAAGTGTAAATCACACCTTTTGGACCCCAACT 2298

Qy 1894 LysProThrValLeuArgSerThrSerThrSerSerProVal-----ArgPro 1909
Db 2299 GCTCAGCTACGATCATGCCACTGCTTCTGGGGCCCTTCCATAAGTCAAGGCCCTGCCA 2358
Qy 1910 AlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyr 1929
Db 2359 GCCTCCGTTTACAACACTGCTGCGGATGCC---CTGGCTGCTCTTGTGTGATGCTGCAGCT 2415
Qy 1930 ProThrLeuMetGluProValLeuLeuProLysGlyAlaProArg----- 1944
Db 2416 TCTGCACCCACAGATGGATGTTTCCAAAAACAAAGAGAGTAAGCATGAAGTCCAGGTTA 2475
Qy 1945 -----ValAlaAlaArgProGluArgProArgProAlaAlaAsn 1956
Db 2476 GAAGAAATTTGAGAAGCAGGTCAGCAGCAGTAGTAGAAGCAGCAGCAGTACAGAGAA 2535
Qy 1957 HisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSer 1976
Db 2536 AACCTGGAGGTGGAGAAGAGATCTGTTCAGTGTGTGTCACCTTCTTTCAGCCCTTCCAAGT 2595
Qy 1977 LysGlySerGluProArgProLeuVal----- 1985
Db 2596 GGCAGGCCCGCCCTCATGCTCAGTAGTGTATTCTGAGGCTGGGAAAGATAAAGGCCCT 2655
Qy 1986 ProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsn-----LeuAla 2003
Db 2656 CTCCCAATCCAGATATGAGGAGAGCTAAGGCCGAGGAGACTACCATTTACTGCA 2715
Qy 2004 ProHisAlaSerProAspProProAlaProProAlaSerAlaSerAspProHisArg 2023
Db 2716 GCTAACTTCATAGACGTGATCATCCCCGGCAAAATGTCCTCGACAGAGTCCGAGGAA 2775
Qy 2024 GluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyr 2043
Db 2776 CGTGGCTTCAAGTTTCAGACTCTTCT-----AGTAGCTTGTCTTCT 2817
Qy 2044 HisGlySerSerTyrSerProGluGlyValGluProValSerProValSerSerProSer 2063
Db 2818 CACAGGTATGAACGGCTAGTAGTGCATTCAGTGAATAGTCCCGCCAGCTCACCTGCA 2877
Qy 2064 LeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGlu 2083
Db 2878 CCACCCCGAGAAAGCCACAGGCTATACGCGAGCATGTTGAAGCAAAATCAAGCAGAA 2937
Qy 2084 GlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaHisLeu 2103
Db 2938 AATGAGTCCACTCGACAGTATGAAGTCCA-----CTG 2970
Qy 2104 ProHisLeuArg-----ProLeuProGluSerGlnProSerSerSerPro 2118
Db 2971 CATCATTTATCGTCCCGAGGAGAAATCACCATCTCCACAGCAACAGCCACCTGCCCCCA 3030
Qy 2119 LeuLeuGlnThr-----AlaProGlyValLysGlyHisGlnArgValValThrLeuAla 2136
Db 3031 TCTTCCAGTACAGAGGAATGGACAGGTGCGCCAGGACCCATCATCATCATCTGCT 3090
Qy 2137 GlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSer 2156
Db 3091 GACCATCATCTGTCAAATTTATCACAAAGATTTTGTAGAAAT-----CAAAGTCCCTCG 3144
Qy 2157 AlaProLeuProAlaProLeuTyrSerPheProGlyAla-----SerCysProValLeu 2174
Db 3145 CAGCTTCTACTTCTACATTTCCAAACTTCCACTCTGCTTGTTCATCCACACCTGTGA--- 3201
Qy 2175 AspleuArgArgProProSerAspLeuTyrLeuProProAsp----- 2189
Db 3202 -----AGAACTAAACCTCAAGCCGTACAGCCAGATCAAGTCTCAGAGTCTGTTG 3255
Qy 2190 HisGlyAlaProAla---ArgGlySerProHisSer-----GluGlyGlyLys 2204
Db 3256 CATCCAGACCAAGTCTCCTAGAGTCTCTCCAGAAATCTTGTGGATAAATCCCGGGGAAG 3315

QY 123 LeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThr 142
Db 649 -----GAAGGGCTGAGGGCT---TCTGCAGATGCTAAGAGATTCAGCATTCGGA 696
QY 143 GlyLysLeuGlu---ProValSerProProSerProProHisThrAspProGluGlu 161
Db 697 GGCACACATGAAGCTCCATCTCTCCAAATTTGCGGCAACCATGTGGAGATGATCAAAAT 756
QY 162 LeuValProProArgLeuSerLysGluLeuLeuLeuLeuLeuLeuLeuLeuLeu 181
Db 757 GCTTACCTTCAAAACTCTCAAAAGGAGAGTAAATACAGAGTATGATCGTGTAGATCGA 816
QY 182 GluLeuThrMetValGlnGlnGlnLeuLeuLeuLysLysLysGlnGlnGlnLeuGlu 201
Db 817 GAAATTCGAAAGTAGAACACGACGATCTTAACTGAAAGGAAACACACACAGCTTGAA 876
QY 202 GluGluAlaLysProProGluProGluLysProValSerProProProLysGluSer 221
Db 877 GAAGAGGACGCTAAACCTCTGAGCCTGAGAGCCGCTGCTCTCTCTCTCTCTCTCT 936
QY 222 LysHisArgSerLeuValGlnGlnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 241
Db 937 AAACACGCGATATGTCGCAATTAATATGATGAGATCGGAAAGAGAGAGAGCT 996
QY 242 HisArgLeuLeuGlyLysGlnValGlnValGlnValGlnValGlnValGlnVal 261
Db 997 CATATAATTTTGAAGTCTTGCCCAAAAGTTGAAGTGCATGCTGATATACCGACCATCA 1056
QY 262 AspThrArgGlnTyrHisGluAsnLeuLysLeuLeuLeuLeuLeuLeuLeuLeu 281
Db 1057 GATACCAAGGTGTACCATGAGAACATCAAGACAAACAGCGTGTAGGAAACAACTCAT 1116
QY 282 LeuTyrPhenylArgArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGlnArg 301
Db 1117 TTATTTTTTAAAGAGAAATATCATGCAAGAAACAAAGGAGCAAAATCTGCCAGCGT 1176
QY 302 TyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgLysGlnLeuAsnProArg 321
Db 1177 TATGATCAGCTCAGTGGAGCATGGGAGAAAGAGTGCAGAGATAGAAATATCTCTCG 1236
QY 322 ArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluLeuArg 341
Db 1237 AGGAAAGCTAAAGAAAGCAAAACCAAGGGAATCTATGAAAGCAGTTTCCAGAAATTCGA 1296
QY 342 LysGlnArgGluLeuGlnGlnArgMetGlnSerArgValGlyGlnArgGlySerGlyLeu 361
Db 1297 AAACAAAGAGAACAGCAAGAAAGATTTTCAG---CGAGTTGGCGAGGAGCTGTCTT 1353
QY 362 SerMetSerAlaAlaArgSerGluHisGluValSerGluLeuLeuLeuLeuLeuLeu 381
Db 1354 TCAGCCACCATGCTAGGAGTGAGCATGAGATTTCTGAAATTAATGATGGCTCTCTGAG 1413
QY 382 GlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValLeuProProMetLeuTyrAsp 401
Db 1414 CAGGAGATAATGAGAAACAAATCGGCAGCTCTCTGTGATTCACCATGATGTTGAT 1473
QY 402 AlaAspGlnGlnArgLysPheLeuAsnMetAsnGlyLeuMetAlaAspProMetLys 421
Db 1474 GCAGACAAAGACGAGTCAAGTTCATTAACATGATGGCTTATGGAGGACCTTATGAA 1533
QY 422 ValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArg 441
Db 1534 GTGTATAAGATAGCCAGTTTATGAAATGTTTGGACTGACCATGAAAGGAGATCTTTAAG 1593
QY 442 GluLysPheMetGlnHisProLysAsnPheGlyLeuLeuAlaSerPheLeuGluArgLys 461
Db 1594 GACAGAGTTTATCCAGCATCCAAACAACTTTCGACTAATTCATCATCATCTTGGAGGAG 1653
QY 462 ThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSer 481
Db 1654 AGTGTTCTCTGATGTGTTTGTATCTATTTAACCAAGAAATGAGATTAATAAGCC 1713
QY 482 LeuValArgArgSerTyr---ArgArgArgGlyLysSerGlnGlnGlnGlnGlnGln 500

Db 1714 CTCGTCAAGGAATATGGGAAACGACAGACGACAAACACGACGACAA----- 1761
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Db 1762 -----ATTGCTGACCTCGCAGAGAGAGAAA 1788
QY 521 AspGluLysGluLysGluLysGluAlaGluLysGluGluGluLys 535
Db 1789 GTAGAGAGAGAGAGAGAGATTAACGACAGAGAGAGAGAGAGAGAGAGAGAG 1833
RESULT 28
BC056862 1741 bp mRNA linear PRI 05-NOV-2003
LOCUS Homo sapiens nuclear receptor co-repressor 1, mRNA (cDNA clone
DEFINITION IMAGE:3904576), partial cds.
ACCESSION BC056862
VERSION BC056862.1 GI:34783946
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1741)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Udwin,T.B., Toshiiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettner,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,K.,
Boutfard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,V.S., Krzywinski,M.I., Skalska,U., Smallus,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
PUBMED 12477932
REFERENCE 2 (bases 1 to 1741)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (25-AUG-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smallus, Michael Smith, Lorraine Spence, Jeff Scott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found

Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasha van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Place: 104 Row: a Column: 3
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis: similarity but not identity to protein.

FEATURES

source

```

1..3025
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="MGC:56355 IMAGE:5604262"
/tissue_type="whole body, adult male"
/clone_lib="Sugano SJD adult male"
/lab_host="DH10B"
/note="Vector: pME18S-FL3"
312..1220

```

CDS

```

/codon_start=1
/product="Similar to nuclear receptor co-repressor 1"
/protein_id="AAH49302.1"
/db_xref="GI:29179651"
/translation="MSSGFPSPQSGFSSEQSPHPSVOYTSSTTHPQEPFVSVDYR
NPLQDOQRPSLSFHPDTPPRRRHGEQFHAISAQEQEALESKRPDIENW
ETHFSRQSDTKVYHDKTNQVMKKLILPKRNHARKQEQIKCRYLELMTSW
EKVWRWENNPRKAKESRTREYREQFPIRKQRQEQERFVGORGTLGSLATIAS
EHEISIIDLSQENNEKQMLQSVPPVNYDSQGRVFINNGLMDDPMPKYYKSR
OFNNVWTEHEKILGRSLCSIPRTSG"

```

ORIGIN

Alignment Scores:	
Pred. No.:	2.03e-25
Score:	1628.50
Percent Similarity:	49.30%
Best Local Similarity:	37.51%
Query Match:	12.32%
DP:	5
Length:	3025
Matches:	404
Conservative:	127
Mismatches:	198
Indels:	350
Gaps:	28

US-09-532-753-5 (1-2517) x BC049302 (1-3025)

[illegible]

31 AlaArgThrHisThrAspValGlyLeuLeuGluTyrGlnHisHisSerArgAspTyrAla 50

DB 405 ACACGCCACCCCGCAGGAGTCTCAGTATCAGACATACCGGCAATCCCCCAGCGGCGC
Qy 51 SerHisLeuSerProGlySerIleIleGlnProGlnArgArgProSerLeuLeuSer 70

459	Db	-----CAGCAGCGTGCAGACCTTCACTACTGTCT	488
71	Ov	GluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSerHisSer	90

Db 489 GAATTCCACCCCTGGACCGAGAGGCCACCGGAG-----CCAGACATGGT 533

Db 534 TATGAGCAGGTTCCACGCCATCTCAGCTCAACAGGAGCAGGAGGCTCTCGAGAGCAAG 593

[illegible]

Qy	127	GlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGlu	144
Db	608	-----	608

Qy	462	ThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGlnAsnTyrLysSer	481
Db	1593	AGTGTCCTCATGTGTGTTTGTAATCTATTTAACCAAGAAATATGAGAATTATAAGCC	1652
Qy	482	LeuValArgArgSerTyr---ArgArgArgGlyLysSerGlnGlnGlnGlnGlnGln	500
Db	1653	CTCCTCAGAAGGAAATATGGGAAACCGACAGGCGAGAAACGAGCAA	1697
Qy	501	GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluLys	520
Db	1698	-----ATTGCTCGACCCCTCGAAGAGAGAAAAA	1724

Qy	521 AspGluLysGluLys	525
	: : : :	
Dβ	1725 GTAAAAA	1739

RESULT 29	BC049302	3025 bp	linear	mRNA	VPT 07-OCT-2003
LOCUS					
DEFINITION	Danio rerio mRNA similar to nuclear receptor co-repressor 1 (cdna clone MGC:56355 IMAGE:5604262), complete cds.				

ACCESSION	BC049302
VERSION	BC049302.1
KEYWORDS	MGC.
SOURCE	Danio rerio (zebrafish)
ORGANISM	Danio rerio

ORGANISM

1. (bases 1 to 3025)

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldi, M.F., Casavani, T.L., Schaeetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shvechenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J. W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Groudman, J., Schmutz, J., Myers, R.M., Butterfield, Y. S., Krzywinski, M.I., Skalska, U., Smalius, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.,

TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

MEDLINE
22388257
PUBMED
12477932
REFERENCE 2 (bases 1 to 3025)
AUTHORS Strausberg, R.

TITLE
Direct Submission
Submitted (24-MAR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk Email: cgabbs-r@mail.nih.gov Tissue Procurement: Dr. Sumio Sugano cDNA Library Preparation: Dr. Sumio Sugano cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada info@cgsc.bc.ca Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Lecticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo

QY 147 ProValSerProProSerProProHisThrAspProGluLeuValProProArg 166
Db 608 ----- 608
QY 167 LeuSerLysGluGluLeuLeuGlnAsnMetAspArgValAspArgGluLeuThrMetVal 186
Db 609 -----AATGTCTCC 617
QY 187 GluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeuGluGluAlaAlaLys 206
Db 618 GAAACACACTTTCACGA ----- 635
QY 207 ProProGluProGluLysProValSerProProLeuGluSerLysHisArgSerLeu 226
Db 635 ----- 635
QY 227 ValGlnIleIleTyAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGlu 246
Db 635 ----- 635
QY 247 GlyLeuGlyProGlnValGluLeuProLeuTyAsnGlnProSerAspThrArgGlnTyr 266
Db 636 -----TCTCAGCCCTCTGATCTAAAGTCTAC 662
QY 267 HisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyPheLysArg 286
Db 663 CATGACAACTTAAGACCAATCAGGTGATGAGGAAGAAGCTGATTCTGTGTTTTCAAGAGG 722
QY 287 ArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGlnArgTyrAspGlnLeuMet 306
Db 723 AGAAATCATGCAGCAACACAGAGGAGCAGAAATTTGTGAGCGGTATGATGAACACTCATG 782
QY 307 GluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysGlu 326
Db 783 ACAGAGTGGGAGAAAAGTTGACGTATGGAAAACACCCACGCCGTAAAGCCAAAGAG 842
QY 327 SerLysValArgGluTyrTyrGluLysGlnPheProGluLeuArgLysGlnArgGluLeu 346
Db 843 AGCGGCACACGAGAATACTACGAGAGGCGAGTTTCTGAAATCCGCAAAACACGCGCGCA 902
QY 347 GlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaAla 366
Db 903 CAGAGCGTTTCCAG---AGAGTTGGCCAGAGGGGACAGGCCCTCTGCGCACCATCGCA 959
QY 367 ArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGlu 386
Db 960 AGAAGTGAAATGATGATTTCTGAGATCATTTGACGGCCCTCTCGAGCAGGAGAACATGAG 1019
QY 387 LysGlnMetArgGlnLeuAlaValIleProProMetLeuTyAspAlaAspGlnGlnArg 406
Db 1020 AAGCAGATGAGCGAGTTGTCCGTCTATTCCTCCCGTGATGATGATTCAGAGCAGCGGAGA 1079
QY 407 IleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyLysAspArg 426
Db 1080 GTAAGTTTATTAATATGAACGGTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1139
QY 427 GlnValMetAsnMetTyrSerGluGlnGluLysGluThrPheArgGluLysPheMetGln 446
Db 1140 CAGTTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1198
QY 447 HisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCys 466
Db 1199 CATCCCAAGAACTTCGGGCTTAATCGCTCTTTCTGAGAGGAGGTGTGTCTGATTGT 1258
QY 467 ValLeuTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 486
Db 1259 GTGCTGTACTACTATTGACTTAAGAGAGTCAAACTATATAAAGCGTGTGTCAGACGAAT 1318
QY 487 TyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 506
Db 1319 TTTGGAAACCGAGGCCCGGCAATCAGCAGATACACAGCCCTTCACAGGAGAGAGACACA 1378
QY 507 GlnGlnGlnGlnProMetProArgSerSerGlnGluLysAspGluLysGluLysGlu 526

Db 1379 GAAGACAAGAACGAAGAGGCAAAATCCGAGAAAACCGAGAGAAAGAGGACGAGGAGAAA 1438
QY 527 LysGluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAspLeuLeu 546
Db 1439 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1486
QY 547 LysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSer 566
Db 1487 AAAGATAAA---GAGAAAGTGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1543
QY 567 LysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIle---ThrArgSer 585
Db 1544 CGTGGCCGAAAGACCGCAACAGCAGCGCGCGCGCAAGAGCGCATCACCACTCGCTCC 1603
QY 586 MetAlaAsn----- 588
Db 1604 ATGCGCAATAATGAGGCTGCATCTGTGTCTGTAAGACCCACCTCCACCACTCGAA 1663
QY 589 -----GluAlaAsnSerGluAlaIle 596
Db 1664 CCAGTTTTTGCAGAGCCAGCACCCCTTCCAAACCTTGAGCGGACACAGAAACAGCTCGA 1723
QY 597 ThrProGlnGlnSerAlaGluLeuAsp-----MetGluLeuAsn 610
Db 1724 GAGCCCGCAAAACCTCTATTGCGAGCCAGCGGACACACAGAGCGCGGTTGAAGCTGGA 1783
QY 611 GluSerSerArgTyrThrGluGluMetGlnThrAlaLysLysGlyLeuLeuGluHis 630
Db 1784 GAGAGTCTCGCTGGAGGAGGAGATGAGATCCCAAAAAGGTCTGTGTTGAACAT 1843
QY 631 GlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLys 650
Db 1844 GCGAGAACTGTCAGCCATTGCTAAGATGATGGAGAGTAAGAGTGAAGCTCAATGTAAG 1903
QY 651 AsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLys 670
Db 1904 AACTTTTACTTCAACCAAGAGCGGATACAACTAGACAGTCTTACTACGCAACACAAG 1963
QY 671 LeuLysMetGluLysGluArgAsnAlaArgGlyLysLysLysAlaProAlaAlaAla 690
Db 1964 -----GGTTCCCGCGGCTTCTGTAATCGAGACCCCTCGAG 2002
QY 691 SerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGlyVal 710
Db 2003 AGTGAT-----CCAGTGCCTGATGACCATGAG----- 2029
QY 711 SerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGlu 730
Db 2029 ----- 2029
QY 731 ValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSer 750
Db 2030 -----GAGAGTCGGTCCA-----GATAACAGTTTCAAGACAGAGAGT 2068
QY 751 IleProSerPro----- 757
Db 2069 GCGCTTCTCCAACCTCAACTGATCCCTCCAAACCTCAGATCACTCACTATACCAGT 2128
QY 758 AlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGly 777
Db 2129 TCAGCAGTGTGATCGCGATCAGAC-----CAGGACGCTCACTGCGCCAAAGGC 2176
QY 778 Pro-----ProProGlyProProThrProThrProThrProThrProThrProThrPro 794
Db 2177 CCGGTGCGCATGCCAAAGCCGAGGCTCTCCATACCGTGAGCTG---AGGTTAAATTT 2233
QY 795 GluProThrProAlaSerGluAlaThrGlyAlaProThrProProAlaProProSer 814
Db 2234 GAGAAGAGTCCAGAGGAGGAGCC----- 2257
QY 815 ProSerAlaProProProValProLysGluGluLysGluGluThrAlaAlaAla 834


```
Qy 1411 sGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgG 1431
Db 62660 TGTGTCAAGGTGTTCACTCATGTGTGAGGATGTTAGGTGTTCACTCATGGCAAGTA 62719
Qy 1431 uGluLeuArg-----HisThrProLueuProLueuAlaProAr 1444
Db 62720 TCACATATACACAGGTGAGGTGTTGGGTATTACACCCATGTGTGAGGTATTCACT 62779
Qy 1444 gProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSe 1464
Db 62780 CGTGAGGATGTGATTGTTCACTCATGTGCAAGAGTGTACATATACACAGSGTGAGG 62839
Qy 1464 r----- 1464
Db 62840 TGTGGGTATTAAACACCCATGTGTGAGGTGTTCACTCATGTGTGAGGATGTTAGGTGTT 62899
Qy 1465 -----ThrThrGlySerLysLysHisAspValArgSerLeuIleGly----- 1478
Db 62900 CACACTCATGGCAAGATATACATATACACAGGTGAGG---GTGTGGGTATTTCACAC 62956
Qy 1479 -----SerProGlyArgThrPheProProVa 1487
Db 62957 CCATAGTGAAGTGTCTAGTATTACACACCCGTGAGTCCA-----GAATGGCCATCACT 63010
Qy 1487 lHisProLeuAspValMetAlaAspAlaArgAlaLeuGlu-----Ar 1501
Db 63011 TCCTAAATTGATGATTCTTAATGTATCTCCCTGCCATAGAGATCATGACCGGGATGTCCT 63070
Qy 1501 gAlaCysTyrGluGlySerLeuLysSerArgProGlyThrAlaSerSerSerglyGlySe 1521
Db 63071 GGTGTGTTTTGAATACAGGAGGAGGAGAGAGCACTCTCTGCCAGCTTTGTTCCAGTGCTC 63130
Qy 1521 rIleAlaArg-----GlyAlaProValIleValProGluLeuGlyLysPr 1536
Db 63131 TTGGGTTCAAGTCTACACATCTTGGGTGTCCTGTCTGTGTGCCAGCATCCCTTCT 63190
Qy 1536 oArgGlnSerProLeuThrTyrGlu-AspHisGlyAlaProPheAlaGlyHisLeu-Pro 1555
Db 63191 GGAGCAGAGTGA---ACCTGGGAATGCCACAGTCTTCACTATTACCCAGTCACCT 63247
Qy 1556 ArgGlySerPro---ValThrMetArgGluProThrProArgLeuGlnGlySerLeu 1574
Db 63248 CGGCCTTGTCCCTTGGTCAGGAGGCTGCGCTGCCCTCCCTCATTT----- 63292
Qy 1575 SerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLys 1594
Db 63293 -----ACC 63295
Qy 1595 SerProHisSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeu 1614
Db 63296 TGCCCTTTCCGTATGTCGCACAGCAGCAGCGCTGCCCC----- 63334
Qy 1615 LeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspPro 1634
Db 63335 -----TCCCACTCCCCCAAGCCTGACACC 63361
Qy 1635 ThrSerIleProArgGlyIle----- 1641
Db 63362 CACTGTGTCGGTCCCGTGTGTCAGTGTGTCAGTCCCGAGCCAGATTGGGTGCTGAGC 63421
Qy 1641 ----- 1641
Db 63422 TNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 63481
Qy 1642 ProLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyr 1661
Db 63482 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 63541
Qy 1662 ProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsn 1681
Db 63542 CCGCACCTGTACCACCTTACCTCATCCGGCGGTACCCCGACACAGCGCGCTGGAGAAC 63601
```

```
Qy 1682 ArgGlnThrIleLeuAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThr 1701
Db 63602 CGGCAGACCATCATCAATGACTATCATCCCTCGCAGCAGATGCACCAACCGCGCCACC 63661
Qy 1702 AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGlySerSerLeuAla 1721
Db 63662 GCCATGCCCCAGCGAGCTGATATGCTGAGGGGCTCTCGCCCCCGCGAGTCTCTCGCTGGCA 63721
Qy 1722 LeuAsnTyrAlaAlaGlyProArg----- 1729
Db 63722 CTCACTACCTCGCGGTGCTCCCGAGGTGAGTGGGTGGGCAGACCACTCCCTCGGTTTG 63781
Qy 1729 ----- 1729
Db 63782 GCCTTATTCCAAAGACATGGCGTCCCTGTGGCTTCGCGAGGCAGCTAGACCTGG 63841
Qy 1729 ----- 1729
Db 63842 TCACCTTGTGGGTACCTTGTGTGAACGGACCTGAGTGGGTGGCTCGGGGTGTGCGTGC 63901
Qy 1729 ----- 1729
Db 63902 TGTGGGTGCTGTTGGCATCTGGTAGGTGAGTGACAGCGTGTGGCTCTCGGCTGCATCC 63961
Qy 1729 ----- 1729
Db 63962 TCAGTGGGTGTGCGTGCATCTGTGTATCTTCTTAGGATACAGGGGCTCAGAGTTTAAA 64021
Qy 1729 ----- 1729
Db 64022 GATCAAAATGTGCCGGGCACTGTTGGCTCATGCTGTAATCCCAGCACTTTGGGGAGGGCG 64081
Qy 1729 ----- 1729
Db 64082 AGGCAGTGTGATACAAGTTCAGAACTTCAGAGACCAGTCTGACCAACATGTGTGAACCCG 64141
Qy 1729 ----- 1729
Db 64142 TCTCTCTAAATAATACAAAAATTAGCCAGGCATGGTGATGGCACCTGTGTGCCAGCTA 64201
Qy 1729 ----- 1729
Db 64202 CTCACTAGGTGAGGCAAGAGAATCACTTGAAACCCAGAGGTGGAGGTTACAGTGAATGG 64261
Qy 1729 ----- 1729
Db 64262 AGATTTTACCATTGCACCTCCAGCTGGGCAACAGGGCAAGACTGTGTCTCGAAAAAATA 64321
Qy 1729 ----- 1729
Db 64322 AAAAGATGAAATGTGAGGCTGTTTGGAGTTTGTTCCTTTTGTGTTAAACAGCCACAG 64381
Qy 1729 ----- 1729
Db 64382 CTGCTTTTGGTGCACAGTTCACGGCCATCTCTCAGAAATGCTTCTCGAAATAACCAAGTT 64441
Qy 1729 ----- 1729
Db 64442 CTAGCTGGGGCTCAGCTGGGAAAAGCTGAAGTCACTTAAGTATTTTGAACAGTGAAGAT 64501
Qy 1729 ----- 1729
Db 64502 TGAATACAGGGAATGGATTGTATAGGCGTCAGAGGCTGAAGGGGCACAGAGGGCTTGAGA 64561
Qy 1729 ----- 1729
Db 64562 TGGGAACCAAGTAGGGCAGCTGCAGGAGATGCCCGGCTCGGGCTTGGGAGAGAAGAGG 64621
Qy 1729 ----- 1729
Db 64622 AGGTGTACCGAGAGAACCTGAGCATTTCAGAAAAAGGTTTCCATGGCCGGTGTCTGGGAGCC 64681
Qy 1729 ----- 1729
```

```
Db 64682 GAGGAGGAGTGCCTGCCACCACTCTGCTGGCTCCAGGAGTGTGTCGCTCTCCAG 64741
Qy 1729 ----- 1729
Db 64742 GAGGGTATCTGGCCGGTGGGAGCCTGGCTCTCTCCCTCTGGCTACAGCCCTGGC 64801
Qy 1730 ----- GlyIleIleAspLeuSerGlnValProHisLeuProValLeuVal 1744
Db 64802 CCAACACCTCCCGAGGCAATCATGACCTGTCCCAAGTGCACACCTGCTGTGCTCGTG 64861
Qy 1745 ProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaThrLeuProThrAla 1764
Db 64862 CCCCCGACACAGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 64921
Qy 1765 ProGlnProPheSerSerArgHisSerSerSerProLeuSer----- 1778
Db 64922 CCCCAGCCCTTCAGCAGCCGCCACAGCAGCAGCTCCGCCACTCTCCCGAGGTAGCGCCTGCC 64981
Qy 1778 ----- 1778
Db 64982 CAGTCTGGGGGTGGGACCCCGGCATCCATGGAGCGGCTGGGGATGGGGCGGCAGA 65041
Qy 1778 ----- 1778
Db 65042 AGCCTGCTCTCTTTCCACCCAGAGACAAAGCCAGGCTCTTCTCGGCCCTGGGGCT 65101
Qy 1779 ----- Pro-GlyGlyProThrHisLeuThrHisProThrThr 1791
Db 65102 GAGTCTCTGGCCCTTTGGGTTTCCTAGGAGGTCCAAACACACTTGACAAAACCAACCCAC 65161
Qy 1791 rSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgGluArgG1 1811
Db 65162 GTCTCTGTCGAGCGGAGGAGCAGCCGGATCGAGAGCGGAGCCGGATCGGGAGCGGA 65221
Qy 1811 uLysSerIleLeuThrSerThrThrThrValGluHisAlaProIleThrParg----- 1828
Db 65222 AAGTTCATCTCTACGCTCCACAGCGGTGGAGCAGCCACCATCTGGAG-ACCTGGTA 65280
Qy 1828 ----- 1828
Db 65281 GGGCATCAGACCCCCACCCCGCTCCGGGACTCTTGTGGCGCCGACAGAGCCCTCCCC 65340
Qy 1829 ----- ProGlyThrGluGlnSerSerGlySerSerGlySerSe 1841
Db 65341 CTGCTGATGCACATGACTGTACACAGGTACAGAGCAGAGCAGCGGAGCAGCGGAGCAG 65400
Qy 1841 rGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHisSe 1861
Db 65401 CGCGGGGGTGGGGCAGCAGCAGCGCCCGCTCCCATCCCATGCCACCCAGCAGCTC 65460
Qy 1861 rProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAenth 1881
Db 65461 GCCCATCTCCCTCGGACCCAGGATGCTCCAGCAGAGACCCAGTGTGCTTACACAC 65520
Qy 1881 rGlyMetLysGlyIleThrAlaValGluProSerLysProThrValLeuArg 1899
Db 65521 AGGCATGAGGGTATCATCCCGTGTGGAGCCAGCAGCCCGCTCTGTGAGG 65575

RESULT 31
LOCUS AK127788
DEFINITION Homo sapiens cDNA FLJ45889 fig, clone OCBF3022166, moderately
similar to Nuclear receptor co-repressor 2.
ACCESSION AK127788
VERSION AK127788.1 GI:34534855
KEYWORDS oligo capping; fig (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,
```

Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,Y., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yanashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahara,K., Masuo,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2914)
Isogai,T. and Yamamoto,J.
Direct Submission
Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES

source

1. .2914
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="OCBF3022166"
/tissue_type="brain"
/clone_lib="OCBF3"
/dev_stage="fetal"
/note="cloning vector: pME18SFL3"
1848. .2273
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC87136.1"
/db_xref="GI:34534856"
/translation="MTEPGHSRNVAVVLLYRDGEOTEPGRMGSKSGNTSQPPAFPSK
LTFESAMVKSKEIKLNTHRNPEYKSGPAGSGMGFPQVGWPAWAAGAAQW
QKPTVHLKLSGPADHLPPGPLPHIWAQFQFLDLLGFP"

CDS

1848. .2273
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC87136.1"
/db_xref="GI:34534856"
/translation="MTEPGHSRNVAVVLLYRDGEOTEPGRMGSKSGNTSQPPAFPSK
LTFESAMVKSKEIKLNTHRNPEYKSGPAGSGMGFPQVGWPAWAAGAAQW
QKPTVHLKLSGPADHLPPGPLPHIWAQFQFLDLLGFP"

ORIGIN

Alignment Scores:

Pred. No.: 1.25e-20 Length: 2914
Score: 1391.00 Matches: 334
Percent Similarity: 49.65% Conservative: 24
Best Local Similarity: 46.32% Mismatches: 94
Query Match: 10.53% Indels: 269
DB: 9 Gaps: 13

US-09-522-753-5 (1-2517) x AK127788 (1-2914)

Qy 1823 HisAlaProIleThrParg---ProGlyThrGluGlnSerSerGlySerSerGly----- 1839
Db 109 TAGCCCCCACCCTGCCCGCTGCTGTGCTGCGGAGTGGGGAGTGGAGCAGCGG 168
Qy 1840 -----SerSerGlyGlyGlyGly 1846
Db 169 TGAAGAGAGAGTGTAGGCGGCGGACCTGCGCTGTGCGAGGAGGAGTGGGTCTGGC 228
Qy 1847 SerSerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArg 1866
Db 229 TCTTGCATCTTGCCTGTGCC-----CAGCCCTGTATCCCCAGA 267
Qy 1867 ThrGlnAspAlaLeu-GlnGlnArgPro-----SerValLeuHi 1879
Db 268 AAAAGGGGAGCCCTCTGCTTGGACCCCTGCTGCGCCCTAGTTCATGCTCTCTCTG 327
Qy 1879 sAsnThrGlyMetLysGlyIleThrAlaValGluProSerLysProThrVal-LeuA 1899
Db 328 TTGACTTGGATGGCCGAGCTATAGCCAGGCGGCGCCCGGGACCCAGGTCATC- 386
Qy 1899 rgSerThrSerThrSerSerProValArgProAlaThrPheProAlaThrHisC 1919

[illegible]

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further information about the clone and the sequencing project is available at <http://www.mips.biochem.mpg.de/proj/cDNA/>.

FEATURES	source	1. .3575	/organism="Homo sapiens"	/mol_type="mRNA"	/db_xref="taxon:9606"	/clone="DKFZp434M075"	/tissue_type="testis"	/clone_lib="434 (synonym: htes3). Vector pSport1; host DH10B; sites NotI + SalI"	/dev_stage="adult"	1. .3575	/gene="DKFZp434M075"	<1. .3145	/gene="DKFZp434M075"	/note="nuclear receptor co-repressor, partly"	/codon_start=2	/product="hypothetical protein"	/protein_id="CAB70854.1"	/db_xref="GI:680844"	/db_xref="SPTREMBL:O9NSZ0"	/translation="SQSAIKHNKSLITGSPKLSRGMPLEIVPENIKVBERGYED VKAGETVRSRHSIVSGSPVSLRSTLHEAPKAQLSPGIVDDTSARETPVSYONTMSRG SPWMNRTSDVTISNKNHRSKTLPTQORESIPAKSPVPGVDPVSHSPDPRHRG STAGEVRSRHLTHLPDAMPFHRALDPAAAYLFOQLSPTPGPGYQYQIYAVENTHQ TILNDYITSOQMVOIPLDPVARGSLPREQPLGLPYPATRGIIIDLTNMPPTILVPHPG TSPPMDRIYIPCTOITPPRPVNSASMPGHTHLAAASAEAREREREREKERERE RIAAASDLVLRGSEOPGRGSHGVRSVPSVQTQETMLQORPSVFOGTNGTSVIT PLDTAQLRIMPLPAGSPISIQGLPASRYNTADALAALVDAASAPQMDVSKTESK HEARLELRSSRAVSSEQQLEKLEVEKRSVQCLYTSAPSPGKQPHSVVYS SLGSHRYETPSDAIEVSPASPPQEQKLTQYQEVKVAQNAENDPTQYEGPLHHY EAGKRGPPKPSYEELRTGKTIITAFNIDIVIITRQIASDKDAREGSQSDSS EAGKRGPPKPSYEELRTGKTIITAFNIDIVIITRQIASDKDAREGSQSDSS RPOQSPSPOQLPPSQAGMGQVPRTHRLITLADHCQIITQDFARNVSSQTPOQ PPTSTFONSALVSTVTRKTSNRYSPESQAOVHQRGSRVSPENIVDKSRGRP GKSPERSHVSSEPEPLSPPOVPVHESKODSLLLSORGAEPAEORNDARSPGISYL PSFTKLENTSPWKSXKQIFKRLNSGGSDMDMAAOPGTEIFNLPAVTTSGSVSS RGHSPADPAGNLGLEDIIRKALMGSPDKVEDHGVMSQPMGVPTGANTSVTSET RREEDPSHSGGVCKPKLISKNSRKSXKPIPGQYLGTERPSSVSSVSHSGDYHRQ TPGWAWEDRPSSTGTOFPYNPLTMRMLSSPTPTPIACAPSAVNQAAPHQONRIWERE PAPLSAQYETLSDD"	3511	/polyA_site	/gene="DKFZp434M075"	ORIGIN	Alignment Scores:	Pred. No.: 3,77e-18	Length: 3575	Matches: 401	Score: 1272.00	Conservative: 149	Percent Similarity: 48.03%	Best Local Similarity: 35.02%	Query Match: 9.63%	Indels: 202	Gaps: 54	US-09-522-753-5 (1-2517) x HSM802410 (1-3575)	QY	1470	LysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisPro	1489	Db	20	AAACAACAATGCAATCTTAATACGGGGCCCTAGCAAACTATCCCTGGGAATGCCTCCG	79	QY	1490	LeuAspValMetAlaAsp---AlaArgAlaLeuGluArgAlaCysTyrGlu-----	1505	Db	80	CTGGAAATGTGCCAGACACATAAAGTGTAGAACGGGGAATAATAGAGATGTGAAA	139	QY	1506	-----GluSerLeuIleSerArgProGlyThrAlaSerSerSerGlySerIleAla	1523	Db	140	GCAGCGAGACCGTGGTTCCTCCGACACACAGTCAGTGGTGAAGCTCTGGCCCTCCGCTTT	199	QY	1524	ArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyr	1543	Db	200	AGTCCACA-----CTGATGAGCTCCCAAGACACACTGAGCCCTGGATTAT	250	QY	1544	GluAspHisGlyAla-----ProPheAlaGlyHisLeuProArgGlySer	1558	Db	251	CATGACACCATGTCACGGAGGAGCCCTGTGATTTATCAAAACACCATCTCCAGAGGTCA	310	QY	1559	ProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSer	1578	Db	311	CCCATG-----ATGAACACAGAACTTCTGATGTTCATATTTCTTAACAAG	355	QY	1579	Ala---SerGlnAspArgLysLeuThrSerProArgGlu-----IleAla	1593	Db	356	TCTACCAATCATGAAGGAATTCGACACTGACCCCTACCGAGGGAAGTATCCACGG	415	QY	1594	LysSerProHisSerThrValProGluHisProHisProHisProHisProHis	1613	Db	416	RAGTCTCCAGTGCCTGGGTGGACCTCGTGAGCCAC-----AGTCCGTTGATCC	469	QY	1614	LeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAsp	1633	Db	470	CATCACAGAGGAGCAGCTCCAGGCGAGGTTCATCGAGCCACCTCCACGACTTGAT	529	QY	1634	ProThrSerIleProArgGlyIleProLeuAsp---AlaAlaAlaAlaTyrTyrLeuPro	1652	Db	530	CCA---GCCATGCCCTTTTACAGGGCTTTGGATCCTCGAGGGCTGTACCTGTTT	586	QY	1653	ArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGly	1672	Db	587	AGACAGCTTTTACCAACTCCAGGTTACCAAGCTAGTATCAGCTTTAC-----	634	QY	1673	TyrProAspThrAlaAlaLeuGluAsn---ArgGlnThrIleIleAsnAspTyrIleThr	1691	Db	635	-----GCAATGGAGAACACACAGACACATCTTAAATGATTACATTACC	679	QY	1692	SerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArg	1711	Db	680	TCACAACATGCAAGTGAAGTTG-----CGTCCAGATGTGCCACA	721	QY	1712	GlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIle	1731	Db	722	GGACTCTCCCAAGAGAGAGCAGCCTGCTGCTCCCATCCCAACGCA---ACGAGAGGAATC	778	QY	1732	IleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrPro	1751	Db	779	ATTGACCTGACCAATATGCTCCA---ACAATTTAGTCTCCATCCAGGGGAACAAGC	835	QY	1752	AlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArg	1771	Db	836	ACTCTCTCCCATGACAGAACTCACTTATATCTCTGGTACACAGATTACTTCCCTCC	895	QY	1772	---HisSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThr	1790	Db	896	CCGTACAACCTCTCTTCCATGTCTCCAGGACACCCCAACACACCTT-----GCAGCT	946	QY	1791	ThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgGluArg	1810	Db	947	CTGCAAGTCTGAGAGGGAACGGGAACGGGAGGAGGAGGAGGAGGAGGAGGAGG	1006	QY	1811	GluLysSerIleLeuThrSerThrThrThrValGluHisAlaProIleTyrArgProGly	1830	Db	1007	-----ATTGCTGCAGCTTCTCTCCGACCTC-----TACCTGCGCCGAGGC	1045	QY	1831	ThrGlnLysSerSerGlySerSerGlySerGlyGlyGlyGlyGlySerSerSerArg	1850	Db	1046	TCAGAACAG-----CCTGGCCGA	1063	QY	1851	ProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAla	1870	Db	1064	CTGGCAGTCATGATATGTCCTCCCTCCCTCCCT---TCAGTAAGAACTCAGGAGACC	1120	QY	1871	---LeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAla	1889	Db	1121	ATGTTGCAACAGACACCGAGTGTTCCTCAAGAACCAATCGAACCACTGTATACACCT	1180	QY	1890	ValGluProSerLysProThrValLeuArgSerThrSerThrSerProVal-----	1907
----------	--------	----------	--------------------------	------------------	-----------------------	-----------------------	-----------------------	----------------------------------------------------------------------------------	--------------------	----------	----------------------	-----------	----------------------	-----------------------------------------------	----------------	---------------------------------	--------------------------	----------------------	----------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------	-------------	----------------------	--------	-------------------	---------------------	--------------	--------------	----------------	-------------------	----------------------------	-------------------------------	--------------------	-------------	----------	-----------------------------------------------	----	------	--------------------------------------------------------------	------	----	----	------------------------------------------------------------	----	----	------	----------------------------------------------------------	------	----	----	---------------------------------------------------------	-----	----	------	----------------------------------------------------------	------	----	-----	--------------------------------------------------------------	-----	----	------	--------------------------------------------------------------	------	----	-----	----------------------------------------------------	-----	----	------	----------------------------------------------------	------	----	-----	-------------------------------------------------------------	-----	----	------	-----------------------------------------------------------	------	----	-----	----------------------------------------------------	-----	----	------	----------------------------------------------------	------	----	-----	---------------------------------------------------------	-----	----	------	--------------------------------------------------------	------	----	-----	-------------------------------------------------------	-----	----	------	--------------------------------------------------------------	------	----	-----	---------------------------------------------------------	-----	----	------	--------------------------------------------------------------	------	----	-----	--------------------------------------------------------	-----	----	------	--------------------------------------------------------------	------	----	-----	------------------------------------------------------	-----	----	------	--------------------------------------------------------------	------	----	-----	--------------------------------------------------	-----	----	------	--------------------------------------------------------------	------	----	-----	----------------------------------------------	-----	----	------	--------------------------------------------------------------	------	----	-----	--------------------------------------------------------------	-----	----	------	--------------------------------------------------------------	------	----	-----	----------------------------------------------------------	-----	----	------	--------------------------------------------------------------	------	----	-----	----------------------------------------------------------	-----	----	------	-----------------------------------------------------------	------	----	-----	-----------------------------------------------------------	-----	----	------	--------------------------------------------------------	------	----	-----	-------------------------------------------------------	------	----	------	--------------------------------------------------------------	------	----	------	----------------------------------------------------	------	----	------	-----------------------------------------------------------	------	----	------	-------------------------	------	----	------	--------------------------------------------------------------	------	----	------	-----------------------------------------------------------	------	----	------	--------------------------------------------------------------	------	----	------	-----------------------------------------------------------	------	----	------	----------------------------------------------------------	------

```

Db      1181  TTTGATCCAACTGCTCAGCTACGAATCATGCTCCACTGCTGCTGGGGCCCTTCAATAAGC 1240
Qy      1908  -----ArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeu 1925
Db      1241  CAAGCCCTGCAGGCTCCGGTTACAACACTGCTCGGATGCC-----CTGGCTGCTCTGTG 1297
Qy      1926  AspGlyValTyrProThrLeuMetGluProValLeuLeuProLys-----Glu 1941
Db      1298  GATGCTGCAGCTTCTGCACCCAGATGATGTGTCCTCAAAACAAAAGAGAGTAAGCATGAA 1357
Qy      1942  AlaProArgValAlaArgProGluArgProArgAlaLeu----- 1954
Db      1358  GCTCCAGGTTAGAGAAATTTGAGAGCAGGTCAGCAGCAGTGTAGTGAACAGCAGCAG 1417
Qy      1955  ----- 1957
Db      1418  CTAGAGCAGAAACCCCTGGAGGTGGAGAGAGATCTGTTCAAGTGTATTATACACTTCTTCA 1477
Qy      1958  AlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSer----- 1974
Db      1478  GCCTTTCCAAGTGCAGCC-----CAGCCTCATTTCTCAGTAGTTAT 1522
Qy      1975  ProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIle 1994
Db      1523  TCTGAGGCTGGGAAGATAAAGGCT-----CCTCCAAATCCAGATATAGGAAGAG 1576
Qy      1995  AlaArgThrProAlaLysAsn-----LeuAlaProHisHisAlaSerProAspPro 2012
Db      1577  CTAAGGACACAGAGAGAGACTACCATTTACTGCTGCTCATCTTCAAGAGTATCATCACC 1636
Qy      2013  AlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSer 2032
Db      1637  CGGCAATTCCTCGGACACAGATCGCAGGGAACGTGCTCTCAAGTTTCAGACTCTTCT 1696
Qy      2033  IleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGly 2052
Db      1697  -----AGTAGCTTATCTTCTCACAGGTATGAACACCTAGCGATGCT 1738
Qy      2053  ValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLys 2072
Db      1739  ATTGAGGTGATAGTCTCGCAGCTCCTCCTCGCCACCCAGGAGAACTGCAGACCTAT 1798
Qy      2073  HisLeuGluLeuLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGly 2092
Db      1799  CAGCCAGAGGTGTTAAGCAATCAAGCGGAATATGATCTTACCAGACATATGAAGGA 1858
Qy      2093  ProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSer 2112
Db      1859  CCA-----TTACATCACTATCGACCA-----CAG 1882
Qy      2113  GlnProSerSerProLeuLeuGlnThrAlaProGly----- 2125
Db      1883  CAGGAATCACCATCTCCCCAACACAGCTGCCCTTCTTCCAGGAGGAGGAATGGGG 1942
Qy      2126  ---ValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThr 2144
Db      1943  CAAGTGCAGGAGCCCATCGCTGATCACACTTCTGATCACATCTGTCAAAATTATCACA 2002
Qy      2145  GlnAspTyrThrArgHisHis-----ProGlnGlnLeuSerAlaProLeu 2159
Db      2003  CAAGATTTCAGAAATCAAGTTTCTTCGCGAGACTCCCCAGCAG-----CCTCCT 2053
Qy      2160  ProAlaProLeuTyrSerPheProGlyAla-----SerCysProValLeuAspLeuArg 2177
Db      2054  ACTTCTACATCCAGAACTACCTTCTGCTTGTGTATCTACACCTGTG-----AGG 2104
Qy      2178  ArgProProSerAspLeuTyrLeuProProAsp-----HisGlyAla 2192
Db      2105  ACTAAACATCAACCGTTACAGCCAGATCCAGGCTCAGTCTGTCCATCATCAAGA 2164
Qy      2193  Pro---AlaArgGlySerProHis-----SerGluGlyGlyLysArgSer 2206

```

RESULT 33
AF303585
LOCUS

AF303585 3997 bp mRNA linear PRI 01-FEB-2003

```

Db      2165  CCAGTTCAAGGGTCTCTCCAGAAAAATCTTGTGGCAAAATCCAGGGGAAGTAGCCCTGGA 2224
Qy      2207  ProGluProAsnLysThrSerValLeuGlyGlyGlyGlyGlyIleGluProValSer 2226
Db      2225  AATCCCGAGAGAGAGTACGTC-----TCTTCGGAGCCCTACGAGCCCATCTCC 2275
Qy      2227  ProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyr 2246
Db      2276  CCACCCAG-----GTTCCGGTTGTGCATGAGAAACAGGACAGCTTCTCTGTCT 2329
Qy      2247  ArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSer 2266
Db      2330  CAGAGGGCGCAGAGCTGCAGAGCAGAGAAATGATCCCGCTCACCAGGAGTATAAGC 2389
Qy      2267  GlnProProAlaPhePheSerLysLeuThrGlnSerAsnSerAlaMetValLysSerLys 2286
Db      2390  TACTTGCTTCATTTCTTCCACAGCTT-----GAAATACATCACCCATGGTTAATCAAG 2446
Qy      2287  LysGlnGluLeuAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIle 2306
Db      2447  AAGCAGGAGATTTTTCGTAAGTTGAATCTCTCTGCTGGAGGTGACTCTGTATATGCGAGCT 2506
Qy      2307  SerGlnProGlyThrGluLeuPheAsnMetProAlaIleThrGlyThrGlyLeuMetThr 2326
Db      2507  GCTCAGCAGGAACTGAGATCTTTAATCTGCCAGCAGTTACTACGTCAAGGCTCAGTTAGC 2566
Qy      2327  TyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIle 2346
Db      2567  TCTAGAGCCATCTTTTGTGATCTCTGCCAGT---AATCTTGGGCTGGAGACATATATC 2623
Qy      2347  ArgLysAlaLeuMetGlyLysTyrAspGlnTyrGlu-----Ser 2360
Db      2624  AGGAAGCTCTCTGGGAAGCTTTGTAGTACAAAAGTTCAGGATCATGGAGTTGTCTATGTC 2683
Qy      2361  ProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAla 2380
Db      2684  CAGCCTATGGAGTA-----GTCCCTGGTACTGCCAACACCTCAGTT----- 2725
Qy      2381  MetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGly 2400
Db      2726  -----GTGACCATGCTGTGAGACACGAGAGAGAGAGAGAGAGGAGGAGGAGGAGGAG 2779
Qy      2401  Gly-----LysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAla 2418
Db      2780  GGAGTTTGCAAAACCAAGCTGATCAGCAAGTCAAAACAGCAGGAAATCTAAGTCTCTATA 2839
Qy      2419  Pro-----GlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGlu 2436
Db      2840  CCTGGCAGGCTACTTTAGGAACGGAACGGCCCTCTTTCAGTCTCTCTGTATACATTCAGAA 2899
Qy      2437  GlyAspCysAsnArgArgThrProLeuThrAsnArgValTyrGluAspArgProSerSer 2456
Db      2900  GGGATTTACATAGCAGACGCCA-----GGTGGGCTGGGAGAGAGAGAGGAGGAGGAGGAG 2953
Qy      2457  AlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMet 2476
Db      2954  ACAGGCTCAACTCAGTTTCTTATACCTCTGATATGCGATGCGGATG-----CTCAGC 3004
Qy      2477  AlaSerProProProGlyLeuProAlaGlySerGly-----ProLeuAlaGlyProHis 2495
Db      3005  AGTACTCCCAACACCGATTGTCATGTGCTCTCGGTGTAACCAAGCAGCTCTCCAC 3064
Qy      2496  -----HisAlaTyrAspGluProLysProLeuLeuCysSerGlnTyrGluThr 2512
Db      3065  CAACAGAACAGGATCTGGAGCGAGAGCTGCCCATCTGCTCTCAGCACAGTACGAGACC 3124
Qy      2513  LeuSerAspSerGlu 2517
Db      3125  CTGTCGATAGTATGAT 3139

```

D	b	585	TAC-----GCAATGGAGAACACAGACAGCATCTTAA	617
Qy	1687	AsnAspTyrIleThrSerGlnGlnMethHisAsnThrAlaThrAlaMetAlaGlnArg	1706	
D	b	618	ANTGATTACATTACTCACAACAGATGAAGTGAACCTTG-----CGT	659
Qy	1707	AlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaala	1726	
D	b	660	CCAGATGTGGCCAGAGGAGCTCTCCCACAAGAGAGCACCGCTGGGTCTCCCATACCACGA	719
Qy	1727	GlyProArgGlyIleLeuAspLeuSerGlnValProHisLeuProValLeuValProPro	1746	
D	b	720	---ACGAGAGGAATCATTGACCTGACCAATATGCCTCCA--ACAATTTAGTCCTCAT	773
Qy	1747	ThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGln	1766	
D	b	774	CCAGGGGGAACAAGCACTCCTCCCATGGACAGAATCACTTATATTCCTGGTACACAGATT	833
Qy	1767	ProPheSerSerArg---HisSerSerSerProLeuSerProGlyGlyProThrHisLeu	1785	
D	b	834	ACTTTCCCTCCCAAGGCCGTACAACTCTGCTCCATGTCTCCAGGACACCCACACACCTT	893
Qy	1786	ThrLysProThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgasp	1805	
D	b	894	-----GCAGCTGCTGCAAGTGTGAGAGGAACGGACGGAGCGGAGGAAGGAG	944
Qy	1806	ArgAspArgGluArgGluLysSerIleLeuThrSerThrThrValGluHisAlaPro	1825	
D	b	945	CGSGAGCGGGAACGG-----ATTGCTGCAGCTTCTCCGACCTC-----	983
Qy	1826	IleTrpArgProGlyThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGly	1845	
D	b	984	TACTGGCGCCAGGCTCAGNACAG-----	1007
Qy	1846	GlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProLysSerPro	1865	
D	b	1008	-----CCTGGCCGACCTGGCAGCTCATGTAATGTTTCGCTCCCTCCCTCCTCAGTA	1058
Qy	1866	ArgThrGlnAspAla---LeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLys	1884	
D	b	1059	AGAACTCAGGAGACCATGTTGTGCAACAGACCACAGTGTGTTTTCCAAGMAACCAATGGAACC	1118
Qy	1885	GlyIleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSer	1904	
D	b	1119	AGTGTATTCACACCTTTGGATCCAACTGCTCAGCTACGAATCATGCGCATCGCTGCTGGG	1178
Qy	1905	SerProVal-----ArgProAlaalaThrPheProProAlaThrHisCysPro	1920	
D	b	1179	GGCCCCCTCAATAAGCAAGCCAGCGCTGCCAGCTCCCGTTACAACTGCTGGGGATGCC---	1235
Qy	1921	LeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLys	1940	
D	b	1236	CTGGCTGCTCTTGTGGATGTGTCAGCTTCTGCACCCAGATGGATGTGTCCAAAACAAAA	1295
Qy	1941	-----GluAlaProArgValAlaArgProGluArgProArgAlaAsp-----	1954	
D	b	1296	GAGAGTAAGCATGAAGCTGCCAGCTTAGAAGAAAATTTGAGACGAGTCAGCAGCGAGTT	1355
Qy	1954	-----	-----	-----
D	b	1356	AGTGNAACAGCAGCAGCTAGACAGAAAACCTGGAGGTGGAGAGAGATCTGTTTCAGTGT	1415
Qy	1955	-----ThrGlyHisAlaPheLeuAlalysProProAlaArgSerGlyLeuGluProAla	1972	
D	b	1416	TTATACACTTCTTCAGCTTTCCCAAGTGGCAAGCCC-----CAGCCCTCAT	1460
Qy	1973	SerSer-----ProSerLysGlySerGluProArgProLeuValProProValSer	1989	
D	b	1461	TCTTCAGTAGTTATTCTGAGCGTGGGAAGATAAAGGCCT-----CCTCCAAAATCC	1514
Qy	1990	GlyHisAlaThrIleAlaArgThrProAlalysAsn-----LeuAlaProHisAla	2007	

Db 1515 AGATATGAGGAAGAGCTTAAGACACAGAGGGAAGACTACCATCTACTGCGAGCTTAACCTTCATA 1574
Qy 2008 SerProAspProProAlaProProAlaSerAlaSerAspProHisArgGluLeuThrGln 2027
Db 1575 GACGTGATCATCACCCGCAAAATTCCTCGACAAAGGATGCGAGGAACGCTGCTCAA 1634
Qy 2028 SerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyHisGlySerSer 2047
Db 1635 AGTTTCAGACTCTTCT-----AGTAGCTTATCTTCTCACAGGTATGAA 1676
Qy 2048 TyrSerProGluGlyValGluProValSerProValSerSerProSerLeuThrHisAsp 2067
Db 1677 ACACCTAGCATGTATGAGGTATAGTCTCCAGCTCCTCCAGCTCCTCCGCCACCCAGAG 1736
Qy 2068 LysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArg 2087
Db 1737 AAATGCGAGACCTATCAGCCAGAGGTTGTTAAGCAAAATCAAGCGGAATGATCCTACC 1796
Qy 2088 ProLysGlnProGlyProValLysLeuGlyGlyGluAlaHisLeuProHisLeuArg 2107
Db 1797 AGACAATATGAAGGACCA-----TTACATCACTATCGA 1829
Qy 2108 ProLeuProGluSerSerSerSerProLeuLeuGlnThrAlaProGly----- 2125
Db 1830 CCA-----CAGCAGGAATACCATCTCCCAACAACAGCTGCCCCCTTCTTCACAG 1880
Qy 2126 -----ValLysGlyHisGlnArgValThrLeuAlaGlnHisIle 2139
Db 1881 GCAGAGGAATGGGGCAAGTCCCGAGGACCATCGGCTGATCACACTTGTGATCATC 1940
Qy 2140 SerGluValIleThrGlnAspTyrThrArgHisHis-----ProGlnGln 2154
Db 1941 TGTCAAATATATCACACAAGATTTTGTCTAGAAATCAAGTTTCTCGCAGACTCCCGCAG 2000
Qy 2155 LeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAla-----SerCysPro 2172
Db 2001 -----CCTCTACTTCTACATTCAGAAATCCAGACTCCTCTCTGTTGGTATCTACACCT 2051
Qy 2173 ValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProAsp----- 2189
Db 2052 GTG-----AGGACTAAACATCAACCGTTACAGCCAGAAATCCCGGCTCAGTCT 2102
Qy 2190 -----HisGlyAlaPro---AlaArgGlySerProHis-----SerGlu 2201
Db 2103 GTCCATCATCAAGACCAAGGTTCAGGCTCTCTCCAGAAATCTTGTGGACAAATCCAGG 2162
Qy 2202 GlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAspGly 2221
Db 2163 GGAAGTAGGCGCTGAAATCCCGAGAGGAGTCACTGTC-----TCTTCGGAGCCC 2213
Qy 2222 IleGluProValSerProGluGlyMetThrGluProGlyHisSerArgSerAlaVal 2241
Db 2214 TACGAGCCCATCTCCCAACCCAG-----GTTCCGGTTGTGCATGAGAAACAGGACAGC 2267
Qy 2242 TyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSer 2261
Db 2268 TTGTGCTCTGTCTCAGAGGGGGCGAGAGCTGCAGAGCAGAGGAATGATGCCCGCTCA 2327
Qy 2262 ProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAla 2281
Db 2328 CCAGGGAGTAAAGCTACTTGCCTTCATCTTCAACCAAGCTT---GAAATATACATCACCC 2384
Qy 2282 MetValLysSerLysGlnGluLeuAsnLysLysLeuAsnThrHisAsnArgAsnGlu 2301
Db 2385 ATGTTAAATCAAGAAGCAGAGATTTTTCGTAAGTTGAACTCCTCTGGTGGAGGTGAC 2444
Qy 2302 ProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGly 2321
Db 2445 TCTGATATGCGAGCTGCTCAGCCAGGAACCTGAGATCTTTAATCTGCCAGCAGTTACTACG 2504
Qy 2322 ThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGly 2341
Db 2505 TCAGGCTCAGTTAGCTAGAGGCCATTCTTTTCTGATCCTGCGCAGT---AATCTTGGG 2561

Qy 2342 LeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGlu----- 2359
Db 2562 CTGGAAGACATTAATCAGGAAGGCTCTCATGGAAGCTTTGATGACAAAGTTTGAGGATCAT 2621
Qy 2360 -----SerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAla 2375
Db 2622 GGAGTTGTCATGTCCTCCAGCCTATGGAGTA-----GTGCTGGTACTGCCAACACC 2672
Qy 2376 SerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThr 2395
Db 2673 TCAGTT-----GTACCAGTGTGAGACACGAGAGAGGAAGGGAGGCCA 2717
Qy 2396 SerProGlyGlyGly-----GlyLysAlaLysValSerGlyArgProSerSerArgLysAla 2414
Db 2718 TCACCTCATTCAGGAGTTTGCACCAACCAAGCTGATCAGCAAGTCAACAGCAGGAAATCT 2777
Qy 2415 LysSerProAlaPro-----GlyLeuAlaSerGlyAspArgProProSerValSerSer 2432
Db 2778 AAGTCTCTTATACCTGGCAAGGCTACTTAGGAACGGAACGGCCCTCTTCAGTCTCTCTCT 2837
Qy 2433 ValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAsp 2452
Db 2838 GTACATTCAGAAGGGGATTACCATAGCAGACGCCA-----GGGTGGGCTCGGAAGAC 2891
Qy 2453 ArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGln 2472
Db 2892 AGGCCCTTTCACAGAGCTCAACTCAGTTTCTTATAACCTCTGACTATGCGGATG--- 2948
Qy 2473 AlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGly---ProLeu 2491
Db 2949 -----CTCAGCAGTACTCCACCAACCAACCGATTGCTGCTCTCGCGTCAACCAA 3002
Qy 2492 AlaGlyProHis-----HisAlaTrpAspGluProLysProLeuLeuCysSer 2508
Db 3003 GCAGTCTCTCACCACACAGACGAGTCTGGAGCGAGAGCGCTGCCCACTGCTCTCAGCA 3062
Qy 2509 GlnTyrGluThrLeuSerAspSerGlu 2517
Db 3063 CAGTACGAGACCTGTGCGATAGTGAT 3089

RESULT 34

AL590153/c

LOCUS

DEFINITION

133947 bp DNA linear VRT 29-JUL-2003
Zebrafish DNA sequence from clone RP71-71M17 in linkage group 8
Contains a novel gene and part of a novel gene for a protein
similar to mouse silencing mediator of retinoic acid and thyroid
hormone receptor (SMRT), complete sequence.

ACCESSION

AL590153

VERSION

AL590153.9

KEYWORDS

HTG; retinoic acid receptor; silencing mediator; SMRT; thyroid

SOURCE

hormone receptor.

ORGANISM

Danio rerio (zebrafish)

REFERENCE

Laird, G.

AUTHORS

Direct Submission

TITLE

Submitted (20-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,

JOURNAL

Cambridge, CB10 1SA, UK. E-mail enquiries:

zfish-help@sanger.ac.uk

On Nov 17, 2001 this sequence version replaced gi:15022264.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information

on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep/FP71-71M17 is
 from a Zebrafish mixed sex BAC library VECTOR: pTARBAC2
 This sequence is the entire insert of clone RP71-71M17 This clone
 was isolated from BAC library RP71-71 created by Pieter deJong and
 provided by Gareth Howell (Wellcome Trust Sanger Institute,
 Cambridge, UK)
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest, except on the rare
 occasion of the clone being a YAC.
 Clone-derived Zebrafish pUC subclones occasionally display
 inconsistency over the length of mononucleotide A/T runs and
 conserved TA repeats. Where this is found the longest good quality
 representation will be submitted.
 Repeat names beginning 'dr' were identified by the Recon repeat
 discovery system (Zhirong Bao and Sean Eddy, submitted), and those
 beginning 'dr' were identified by Rick Waterman (Stephen Johnson
 lab, WashU). For further information see
http://www.sanger.ac.uk/Projects/D_rexio/fishmask.shtml
 ----- Genome Center

Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: zfish-help@sanger.ac.uk

FEATURES
 source

Location/Qualifiers
 1. 133947
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="RP71-71M17"
 /clone_lib="RP71-71"
 complement(245..378)

repeat_region /note="TDR20 repeat: matches 109. .242 of consensus"
 repeat_region /note="complement(380..481)
 /note="Dr000409 repeat: matches 1. .102 of consensus"
 repeat_region 499. .872
 /note="HE1_DR1 repeat: matches 2. .384 of consensus"
 repeat_region 873. .882
 /note="2.5 copies 4 mer GAAT 20% conserved"
 repeat_region 1035. .1046
 /note="6.0 copies 2 mer TG 24% conserved"
 repeat_region 1049. .1065
 /note="8.5 copies 2 mer AG 34% conserved"
 repeat_region 1194. .1204
 /note="2.8 copies 4 mer GTTT 22% conserved"
 repeat_region 1287. .1301
 /note="15.0 copies 1 mer T 30% conserved"
 repeat_region 1393. .1405
 /note="2.2 copies 6 mer TCTCAT 26% conserved"
 repeat_region 1404. .1420
 /note="2.1 copies 8 mer TTGATGAT 34% conserved"
 repeat_region 1448. .1457
 /note="10.0 copies 1 mer A 20% conserved"
 repeat_region complement(1688..2225)
 /note="Dr000393 repeat: matches 3. .549 of consensus"
 repeat_region 2252. .2412
 /note="dr759 repeat: matches 2. .131 of consensus"
 repeat_region 2755. .2767
 /note="2.6 copies 5 mer TTGA 26% conserved"
 repeat_region complement(2841..3117)
 /note="dr1203 repeat: matches 40. .310 of consensus"
 repeat_region 3141. .4315
 /note="Dr000241 repeat: matches 3. .1153 of consensus"
 repeat_region complement(4343..4836)
 /note="dr994 repeat: matches 2. .493 of consensus"
 repeat_region 4912. .4925
 /note="14.0 copies 1 mer A 28% conserved"

repeat_region 5050. .5060
 /note="2.2 copies 5 mer TGA 22% conserved"
 repeat_region complement(5528..5702)
 /note="Dr000365 repeat: matches 5. .182 of consensus"
 repeat_region 6032. .6046
 /note="7.5 copies 2 mer AC 30% conserved"
 repeat_region 6351. .6360
 /note="2.0 copies 5 mer TTA 20% conserved"
 repeat_region 6430. .6447
 /note="3.0 copies 6 mer TAAAA 36% conserved"
 repeat_region complement(6463..6739)
 /note="Dr000085 repeat: matches 217. .518 of consensus"
 repeat_region complement(6736..6845)
 /note="dr1076 repeat: matches 1. .118 of consensus"
 repeat_region complement(6852..7008)
 /note="Dr000325 repeat: matches 89. .248 of consensus"
 repeat_region complement(7038..7356)
 /note="Dr000208 repeat: matches 441. .745 of consensus"
 repeat_region 7362. .7385
 /note="2.2 copies 11 mer TTTAAAGTCA 39% conserved"
 repeat_region 7420. .7440
 /note="2.1 copies 10 mer TTTGTACATG 33% conserved"
 repeat_region 7457. .7474
 /note="2.6 copies 7 mer AATATT 27% conserved"
 repeat_region 7468. .7481
 /note="2.8 copies 5 mer ATTTA 28% conserved"
 repeat_region 7669. .7678
 /note="5.0 copies 2 mer AG 20% conserved"
 repeat_region 7772. .7782
 /note="5.5 copies 2 mer TG 22% conserved"
 repeat_region 7899. .7918
 /note="3.3 copies 6 mer CTTTCT 22% conserved"
 repeat_region 7904. .7918
 /note="7.5 copies 2 mer CT 21% conserved"
 repeat_region 7944. .7955
 /note="2.0 copies 6 mer CACAGT 24% conserved"
 repeat_region 8150. .8169
 /note="2.0 copies 10 mer TCATGATTTA 40% conserved"
 repeat_region 8349. .8363
 /note="2.5 copies 6 mer TTGTAT 21% conserved"
 repeat_region complement(8394..8560)
 /note="HE1_DR1 repeat: matches 206. .374 of consensus"
 repeat_region complement(8530..8606)
 /note="Dr000009 repeat: matches 771. .848 of consensus"
 repeat_region complement(8549..8614)
 /note="DNA9TA1_DR repeat: matches 1196. .1261 of consensus"
 repeat_region complement(8629..8681)
 /note="HE1_DR1 repeat: matches 334. .384 of consensus"
 repeat_region 8682. .8704
 /note="3.3 copies 7 mer TATTATT 37% conserved"
 repeat_region 8684. .8704
 /note="5.2 copies 4 mer TTAT 35% conserved"
 repeat_region 8710. .8721
 /note="Dr000377 repeat: matches 153. .231 of consensus"
 repeat_region complement(8722..8781)
 /note="HE1_DR1 repeat: matches 47. .115 of consensus"
 repeat_region 8876. .8885
 /note="2.0 copies 5 mer ATATA 20% conserved"
 repeat_region 8951. .8990
 /note="2.9 copies 14 mer AATTATGATCAATC 46% conserved"
 repeat_region 9147. .9156
 /note="2.5 copies 4 mer GTGT 20% conserved"
 repeat_region complement(9327..9377)
 /note="LOOPERN2_DR repeat: matches 845. .895 of consensus"
 repeat_region 9622. .9642
 /note="10.5 copies 2 mer CA 24% conserved"
 repeat_region 9794. .9805
 /note="6.0 copies 2 mer TG 24% conserved"
 repeat_region 9820. .9829
 /note="5.0 copies 2 mer GA 20% conserved"
 repeat_region 10116. .10565
 /note="Dr000322 repeat: matches 5. .519 of consensus"
 repeat_region 10676. .10806

Qy	1581	GlnAspArgLysLeuThrSerThrProArgGluile-----AlaLysSerProHisSer	1598
Db	78408	CAGAGAAAGGCGCCACCCACCCCAAGAGAGATGGCTCCACTAAGTCTCCA---TTG	78352
Qy	1599	ThrValProGluHisProHisProLysSerProLysLeuLeuArgGlyVal	1618
Db	78351	GCAGTGGGTGAGCACACTGCAAAAT-----TTAGCCCTTCGAGAGGATCCTT---GGGTG	78301
Qy	1619	SerGlyValAspLeuThrArgSerHisIlePro---LeuAlaPheAspProThrSerIle	1637
Db	78300	---GGACCGGACATGATACCGTAGACAGAGTCTCATTTAACATTTGATGCTGCAGCTCG	78244
Qy	1638	ProArgGlyIleProLeuAspAlaAla-----	1646
Db	78243	CCACGGGTATCCCGATTGACTCAGG-TATTCCACATATTTGAGATTAGCAGAAAATGT	78185
Qy	1646	-----	1646
Db	78184	CAGTGTGCAATTTTCCATGATCATTAATATACATTGGAACATTAATAATTTTCATGTATT	78125
Qy	1647	-----AlaAlaTyrTyrLeuProArgHis	1654
Db	78124	ATGTTATGATCATCTCAGTTTGTCTTTGCTTTCCACACGCTATTACCTGCTCTCAC	78065
Qy	1655	LeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuileArgGly---Tyr	1673
Db	78064	CTGGCTCTGGCCCTGGCTACCCACACCCCTATCCACCTACCTGATCAGAGCTGTTTC	78005
Qy	1674	ProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGln	1693
Db	78004	CCTGAACAGAGCCCTGGAAACAGACAAACGGTTTTCATGACTACATTAACCTCTCAG	77945
Qy	1694	GlnMetHisHisAsnThrAlaThrAla-----MetAlaGlnArgAlaAspMetLeuArg	1711
Db	77944	CAGATGCACCACTGGCTGCTGCGCTGCAATGCTGCCAGAGACCTGACCTGCTCAGA	77885
Qy	1712	GlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGly---	1730
Db	77884	GGCTTCACACCTAGAGACAGTCTTAACTTAGCTTATTCACCCAGCCCTAGAGGTGAG	77825
Qy	1731	-----Ileile	1732
Db	77824	CCTTCTGAAAGCGTCAAACTTACATTAACATAGTATTCACTTCTTATTGTTTTTG	77765
Qy	1733	AspLeuSerGlnValProHis-----	1739
Db	77764	TCTGTCACTCAAGTCAGTCATCAAGCAAGATTTCAACACAGGAATAGACACTTTACAAA	77705
Qy	1740	---LeuProValLeuValPro-----	1745
Db	77704	AGATTGAGCATTTATATATACATTTTATTTAAATTAACAATGGTATTAATTTTAGA	77645
Qy	1746	-----ProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyr	1760
Db	77644	GGTACATTTTATATGTTTTCTTCAGGAAACATCTGCACTCCCATGATGATCATCATAT	77585
Qy	1761	LeuPro---ThrAlaProGlnProPheSerSerArgHisSerSerProLeuSerPro	1779
Db	77584	ATCCAGGAACCTTCTCCAGGTTTCCCTAGCGGAGCTCAACACCTCTCTCATCTCACCA	77525
Qy	1780	Gly-----	1780
Db	77524	GGTACAGATGATGCTTAATTTCAACATATTTTGTGAACAAGTTTCAGTTCCCTCACC	77465
Qy	1781	-----GlyProThrHisLeuThrLysProThrThrTh	1791
Db	77464	TTAATTAATGCTTATCTTCTTAACAGTAGGAGCTCTCATATGAGCAAGATCCAAAGTGG	77405
Qy	1791	rSerSerSer-----	1794
Db	77404	TTCATCATCTGCTGAAGAGAAAGAGAACCGGAGAGAGAGAGAGAGAGAGAGAGAGA	77345
Qy	1795	-----GluArgGluArgAspArgAspArgGluArgAspArgGluArgGlu	1811
Db	77344	CGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	77285
Qy	1811	uLys-----SerIleLeuThrSerThrThrThrValGluHi	1823
Db	77284	GAAAGATCGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	77225
Qy	1823	SalAProIleThrArgProGly-----	1830
Db	77224	CGCACCTATATGGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	77165
Qy	1830	-----	1830
Db	77164	ATCATTCTTAAATATCTGATAGTTAAATTTATTGGAACATATGATTTCTTTTAAATGTTT	77105
Qy	1830	-----	1830
Db	77104	CTTTTTTTCATATTTTACATATTTTCACTAATTTTCACTAATTTTCTGGTAATAGCTTAAGTTAA	77045
Qy	1831	-----ThrGluGlnSerS	1835
Db	77044	AACATTGTTGATAGTTAAATTTAATTAATGAGTTGAAAGTGAACCCGAAAGCATTT	76985
Qy	1835	erGlySerSer-----	1838
Db	76984	CTTCTTCAAGCAGTCAATTTGTAGATTAGCTTTTAGTCACCTTCGGTGAATCTTTTAAATG	76925
Qy	1839	-----GlySerSerGlyGlyGly	1845
Db	76924	CATCTCAGTATTGCTAAAAATCTGAACCTTTGTTCTGAGCAGGAGCACCTGAACAGATGA	76865
Qy	1845	lyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProLysSerP	1865
Db	76864	CGGAGCAGAGTCTCGTCCCTCTCTCATCTTACAGTCACACAG---TCCCGCTCTCGC	76808
Qy	1865	roArgThrGlnAspAlaLeuGlnArgProSerValLeuHisAsnThrGlyMetLysG	1885
Db	76807	CCCGTCCCCAGACAGCTTGCAGACAGACCAAGCGTCTGCACCAACACCACTCCAAGA	76748
Qy	1885	lyIleileThrAlaValGluProSerLysProThrValLeuArg-----S	1900
Db	76747	GCCTGTCAAACTCT---GAGCACAACTCTTCTGTGCTGCGGTGAGTCTTCCCACCTC	76691
Qy	1900	erThrSerThrSerSerProValArgProAlaAlaThr-----	1912
Db	76690	AAACTGCTGTCAAGCTCCATCTA-GTGAGATCAACATGTTTAAAGCTGGCTCTTAATTG	76632
Qy	1913	-----PheProProAlaThrHisCysProLeu-----GlyGlyT	1924
Db	76631	TTTCTGCTTTTCTTCTTCAGACCCAGGCTCTGCTCCTACCAGGCTTCAGTGGCA	76572
Qy	1924	hrLeuAspGly---ValTyrPro-----ThrLeuMetGluProValLeuLeuPro	1940
Db	76571	CTTCCAGAGTCTGGTTTACCCCGTCTGAGGTTTACCCATCTGGAACAGAGCCCAAGTCTAA	76512
Qy	1940	ysGlu-----AlaProArgValAlaArgProGluArgProArgAlaAspT	1955
Db	76511	AGACTTGAGCAGAGATGGGGGCAAAAGCCATGGCGAGGAGGTCTCCCAACACCAAGA	76452
Qy	1955	hrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluPro---AlaSerS	1974
Db	76451	CCTTTTCTCTCTTAAATCTGACTCCAGCTCCGAGTCTGCTGGTGGGTGATACATA	76392
Qy	1974	erProSerLysGlySerGluPro---ArgProLeuValProProValSerGlyHisAlaT	1993
Db	76391	TCCCTCTCCATACACAGCCCTCTGTGCCCACCTGCCCAAGTCGAGGGC-----	76339
Qy	1993	hrIleAlaArgThrProAla-----LysAsnLeuAlaProHisHisAlaSerProAspP	2011
Db	76338	-----ACCCCTTGTGGCGCCGAAAGTGGCGGCGGCGGAGTTCATCCATC	76290
Qy	2011	roProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProp	2031

```
Db 76289 CCGTGGCGGGACAGCGCAGCAGCGG -TCAAGTAGGAGAAAGACTCAAAACAAAGTGA 76231
Qy 2031 heSerileGlnGluLeuArgSerLeuGlyTyrHisGlySerTyrSerProG 2051
Db 76230 TGTCATACAGAACAGCACTTCGACACTCGGT-----AAGACCAACATGA 76183
Qy 2051 luGlyValGluProValSerProValSerProSerLeuThrHisAspLysGlyLeuP 2071
Db 76182 CAGCGGCGCAACTTCATAAACCGGATATCATGTCATCAAAATTCCTGTGATCGCGGATGC 76123
Qy 2071 roLysHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeu-ArgProLysGln 2090
Db 76122 CAGAG-----ACTGGTGGCGCTCGCGACCAACGCGCA 76093
Qy 2091 ProGlyProValLysLeuGlyGluAlaAlaHisLeuProHisLeuArgProLeuPro 2110
Db 76092 CCTGTGTGTGTAAGATGCTCTGGGACCTTGGGCGGTTA----- 76054
Qy 2111 GluSerGlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGln 2130
Db 76053 -----ACGCCACATCAAGCCCTCCCTGCCCATTCCTCTGTTCTTACCAAGTCCACC 76000
Qy 2131 ArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHis 2150
Db 75999 AAAACCCCTCTCCCTCCCTCTCCCTC----- 75973
Qy 2151 HisProGlnGlnLeuSerAlaProLeu-----ProAlaProLeuTyrSerPhe 2166
Db 75972 ---CCTCCTCCCTCATCATTCCTCTGCCCATCTCGCCATCTCGCCCTTACGTCGTGG 75916
Qy 2167 ProGlyAlaSerCysProValLeuAspLeuArg 2178
Db 75915 CCTAGTCACGCTGTGTGTGCTTGGCATTAATGAGA 75880

RESULT 35
AX677743
LOCUS AX677743 650 bp DNA linear PAT 27-MAR-2003
DEFINITION Sequence 521 from Patent WO02086122.
ACCESSION AX677743
VERSION AX677743.1 GI:29335148
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Legrain, P. and Daviet, L.
TITLE Protein-protein interactions in adipocytes
JOURNAL Patent: WO 02086122-A 521 31-OCT-2002;
Hybridgenics (FR)
FEATURES
source
location/Qualifiers
1..650
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 1.57e-15 Length: 650
Score: 1113.00 Matches: 216
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.42% Indels: 0
DB: 6 Gaps: 0

US-09-522-753-5 (1-2517) x AX677743 (1-650)
Qy 203 GluAlaAlaLysProGluProGluLysProValSerProProLleGluSerLys 222
Db 2 GAGCTCCCAAGCCCGCCGAGCTGAGAGCCCGTGTACCGCCGCCCATCGAGTCGAAG 61
Qy 223 HisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaHis 242
```

```
Db 62 CACCGCAGCCTGTGTGCAGATCATCTACGACGAGAACCGAAGAGGTGAAGCTGCACAT 121
Qy 243 ArgIleLeuGlyLeuGlyProGlnValGluLeuProLeuTyrAenGlnProSerAsp 262
Db 122 CGGATTCGGAAGCCCTGGGGCCCGCAGGTGAGCTGCCGTGTACACACCCCTCCGAC 181
Qy 263 ThrArgGlnTyrHisGluAenIleLysIleAenGlnAlaMetArgLysLysLeuIleLeu 282
Db 182 ACCCGGCAGTATCATGAGAACATCAANAATAAACACGCGGATCGGAAGAACTAATCTTG 241
Qy 283 TyrPheLysArgArgAenHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyr 302
Db 242 TACTTCAAGAGGAGGAATCAGCTCGGAACAATGGAAGCAGAAAGTTCTGCGAGCGTAT 301
Qy 303 AspGlnLeuMetGluAlaLeuGluLysValGluArgIleGluAenAsnProArgArg 322
Db 302 GACGAGCTCATGGAGGCTTTGGAAGAAAGGTGAGCGCATCGAAGAACACCCGCGCGG 361
Qy 323 ArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLys 342
Db 362 CGGCGCAAGGAGAGCAGAGGTGGCGAGTACTACGAAAGCAGTTCCCTGAGATCGCAAG 421
Qy 343 GlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSer 362
Db 422 CAGCGCGAGCTGCAGGAGCGCATGTCAGCAGGAGGTGGCGCGGCGGCGGTGTGCC 481
Qy 363 MetSerAlaAlaArgSerGluHisGluValSerGluIleIleAepGlyLeuSerGluGln 382
Db 482 ATGTGCGGCCCGCCGACGCGAGCAGAGGTGTACAGATCATCGATGGCTCTCAGAGCAG 541
Qy 383 GluAenLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAla 402
Db 542 GAGAACCTGGAGAAGCAGATGGCGCAGCTGGCCGTGATCCGCGCCATGCTGTACGAGCT 601
Qy 403 AspGlnGlnArgLysPheIleAenMetAenGlyLeuMetAlaAsp 418
Db 602 GACCAGCAGCGCATCAAGTTTCATCAACATGAACGCGCTTATGGCCGAC 649

RESULT 36
AX6775223
LOCUS AX6775223 11296 bp mRNA linear INV 06-DEC-1999
DEFINITION Drosophila melanogaster SANT domain protein SMRTER (Smrter) mRNA,
complete cds.
ACCESSION AF175223
VERSION AF175223.1 GI:5815244
KEYWORDS
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
REFERENCE 1
AUTHORS Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
TITLE Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
JOURNAL Ephydroidea; Drosophilidae; Drosophila.
MEDLINE 1 (bases 1 to 11296)
PUBMED Tsai, C.C., Kao, H.Y., Yao, T.P., McKeown, M. and Evans, R.M.
AUTHORS SMRTER, a Drosophila nuclear receptor coregulator, reveals that
TITLE ECR-mediated repression is critical for development
JOURNAL Mol. Cell 4 (2), 175-186 (1999)
FEATURES
source
location/Qualifiers
1..11296
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
1..11296
/gene="Smrter"
283..10692
/gene="Smrter"
```


QY	773	-----LeuGlyAlaAspGlyP	778
Db	4652	CCAGCGGACAAACAGCCACAGCTTCGGCGATCTCGTTCGGCTATTGGCGGATGCA	4711
QY	778	roProProGly-----	781
Db	4712	GTGGCAACTCGATGTCATGCAATGCAACAGAGATTCTGCCCTCGATGCGAAAGACAATGG	4771
QY	782	-----ProProThrProArgAThrSerA	791
Db	4772	CCAGTGTCTTGTGTGCAAGCGGAGCGGTCTCCCGCAGCCCGTCCACTAAAGAGGGGC	4831
QY	791	rgAla-----ProileluProThrProLaser-----	800
Db	4832	GTGGCCAGCAGTAGTGGCATTCGGAGCAACAAATACCGCGGTGCCCGCTCTGCAATA	4891
QY	801	-----GluAlaThrGlyAlaProThrProProAlaProProSerProSerAlaProP	819
Db	4892	GCTGTCAATGCAATCGGTGCGAGCGCTATCCAAATTGTCGCTACCCACATGCCCAA	4951
QY	819	roProVal-----ValProLysGluGluLysGluGluGluT	831
Db	4952	ATCCCAAGGATCGAGCCAGGCTCTGGGAACATTCCCTCCCGTCTCTCGAA-----	5004
QY	831	hrAlaAlaAlaProProValGlu-----GlyGluGluGluLysProProAlaA	848
Db	5005	-----CTGGCCCGGAGGTGGTGTGATCGGTGATGGCGGAATTTCAAAATACCCACCGATG	5059
QY	848	la-----	848
Db	5060	CAACACGCTGTCTTGGCGCTGCCTAATGCGCATTCGTGCAAGCTGATCCGAGCTGA	5119
QY	849	-----GluGluLeuAlaV	853
Db	5120	ATCTCACGATGGTCCAGTGGCGTCCGCGAGCGGAGTGGCGGCGATGACGCGATG	5179
QY	853	alAspThrGlyLysAlaGlu-----GluProValLysSerGluCysThrGluGluAlaG	871
Db	5180	TCAGTACCTCGTGTGTGATGAACAGAGAGCGGCGGATCCGATACCGTTCCTGCA-G	5238
QY	871	luGluGly-----ProAlaLysGlyLysAspAlaGluAlaAlaG	884
Db	5239	ATCCCGAGAAATTCACGCGGCACAAAGTCCCTTACGATGGTCAAAACAGCAACAGACAG	5298
QY	884	luAlaThrAlaGluGlyAlaLeuLysAla-GluLysLysGlu-----	897
Db	5299	CAGCAACAGCAGCAGCAACAAACAGCAGCAGCAGCAACAGCAACAACTATCG	5358
QY	897	-----	897
Db	5359	CAGCCACAGCCACCTCCGCGCAGCTCCGCGAGCAGAGAGGGTTCGTCTGGAGAGTGGT	5418
QY	898	-----GlyGlySerGly	901
Db	5419	GATCAGGGTACCACTGATTATACACCAACCCGAATGACGAGCAAGTCCGGATCGGGC	5478
QY	902	ArgAlaThrAla-----	906
Db	5479	GGTGGCAAAACAGCGGGCGACAATGAACGCTTACTGCCACCCGCGTGGCGGACCCC	5538
QY	907	LysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGluVal	926
Db	5539	ALAGCAAAAGACAGCAGGAGNAATACGATTCTTCGGCCACGAAACCGCGGATGAGGAG	5598
QY	927	AspGlu-----	928
Db	5599	AACGAAAACTCGCGCGCCAATCGTCAGAGTCCCAAGGTACTATTCCATGGCCATGGACAT	5658
QY	929	AlaGluGlyGlyAspLysSerArgLeuLeuSerProArgPro-----SerLeuLeu	945
Db	5659	GGACATGGTGGTTCATCCCAATATGTGGTGTCTGCAGCCCGCGTGCACCAATATGGG	5718
QY	946	ThrProThrGlyAspProArgAlaAsnAlaSerProGlnLys-----ProLeuAsp	962
Db	5719	ACGGTGGTGGTGTTCACCCAGGTGGAGTGGTGGGCAACAGGTCAACGACCGGATTAGT	5778
QY	963	LeuLysGlnLeuLysGlnArgAlaAlaAlaAlaProProleGln-----ValThrLys	980
Db	5779	ATG-----CGACGCGAGCGGTCAACAATGTTCCAGGATTCGGTATTCTCG	5823
QY	981	ValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProPro	1000
Db	5824	GTAATTGAG-----CGTTCGTGAAGCACACAGGTCGCAACCCGAAG	5865
QY	1001	ProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArg	1020
Db	5866	GGCGACAGGTCAGCAGCAGGG-TCAGGGGCAAGGACAGGGCCAGGTCAGGTCAGAC	5924
QY	1021	GlyLysSerArgSer-----ProAlaProProAlaAspLysGluAlaPheAlaAlaGluAla	1039
Db	5925	TCCAGGTCAAAAGCAGTCCCATCCAGCAGCAACAGCAGCAGCAGCAGCAATCTGC	5984
QY	1040	GlnLysLeuProGlyAspProCysTrpThrSerGlyLeuProPheProValPro-Pr	1059
Db	5985	CAATAATCTGGAACCAAGGAGCTTACCATTTGAAGGGAATACCG-----CCAGGA	6035
QY	1059	oArgGluValLysAlaSerProHis-----AlaProAspProSe	1073
Db	6036	TCCGGGTATCTGAAGCAGCAGCAGCAGCAGCTGGCGGAGCAGCCGCCACCTC	6095
QY	1073	rlAlaPheSerTyAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaAr	1093
Db	6096	GGCTCCGGTAGTTTGGCGCATGGC-----ACCTCCGTACA	6131
QY	1093	gProValLeuProArgProProThrIleSerAsnProProProLeuLysSerSerAlaLys	1113
Db	6132	GAAGTAAACACCCGCGCAGCTGCAGTGCACACACCGCCACCG-----GC	6176
QY	1113	shisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLe	1133
Db	6177	TCATCCA-----CTGACACCCACGAGTTCGATGGTGGGCGGACCAACACGACCGCA	6233
QY	1133	uHisValProTySerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuPr	1153
Db	6234	TAGCTTAGCCACTTATCGTGGTCAATTCGCACATGGCATGGTGGGAATTTGGTCATCC	6293
QY	1153	oLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluLeuSe	1173
Db	6294	GGGCCGATGGCCACCAAGTTCCGCGAGGAATCCGTTGGAGACAGGCGACCAACAC	6353
QY	1173	rPro-----ArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGl	1191
Db	6354	GCCGTTGTCAAGAGCTCTAGTGGATCTAGAGAGCGCGCGCGGCGGCGTCCAGTCA	6413
QY	1191	uAlaSerValLeu-----ArgGlyThrAlaLeuGlySerValProGlySerIleThrLy	1210
Db	6414	CTGCACAGCCACACCGCGGAGACCAATATCTACATGTGCCAGTGGCGCATCCGCAACG	6473
QY	1210	sGlyIleProSerThrArgValProSerAspSerAlaIle-----ThrTyArgGlySe	1228
Db	6474	TGTTATACCA-----CCGTCGACGACTTCGTTACATCCCGGCCCATCCGTCGCA	6524
QY	1228	rlIleThrHisGlyThrProAlaAspValLeuTyLysGlyThrIleThrArgIle---Il	1247
Db	6525	TACACAGACCCCGCCCATCCGCGAGCATTCGTCGATGGTTCAGCATACGCGAGCTGAAGT	6584
QY	1247	eGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHi	1267
Db	6585	GCCGAAACCGGAGCGCAACATTTGACATCAAGAACCGCGCGCGCATGTGCA	6644
QY	1267	sValIleTyGluGlyLysLysGlyHisValLeuSerTyGluGlyGlyMetSerValTh	1287
Db	6645	CTCCCGCAGCACCGGA-----GCCGAGGATCTAGCTC	6677
QY	1287	rGlnCysSerLysGluAspGlyArgSerSerSer-----GlyProProHisGl	1303

6678	Db	GTCTGGCTCCGATCGGGTGACCGTCTCTCATCGATCGCATCATGCGCCACACCGCC	6737
1303	Qy	uThrAlaAlaProLys-----ArgThrTyrAspMetMetGluGlyArgValGlyAr	1320
6738	Db	GACCATGTGCGATGAAGCACATTTGTATGCTCTGGTGGCATGTATCGCGGTGCACACCGTAAC	6797
1320	Qy	gAlaIleSerSerIle-----SerLeuGluGlyLe	1330
6798	Db	TGTTCCATCGCTGCTGCCACCCAGTTTGGTATCTGTATCTCCACACGCGAGTGTCAAGTCCAT	6857
1330	Qy	uMetGlyArgAlaIleProProGluArgHisSerProHisIleSerLeuGluHisIle	1350
6858	Db	TGGCGGGCGGTGGTGGTGGCC-----	6879
1350	Qy	sIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGln-----	1367
6880	Db	-----GGTGTGCTGCCGGTGGCTGGTATTGCACTATATCTACACCCAGTACC	6929
1368	Qy	-----GluAspTyrLeuArgArgGluAlaLysLeuLysArgGluGlyThrPr	1384
6930	Db	AGTCCCGTATCCCAATTTCCATTTCCCGGTCAAGGCCAGTGTGCGCGCAAGAGTGCACAACC	6989
1384	Qy	oProPro-----ProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLe	1402
6990	Db	GCCACCGGCACAGCCGCCATCCGCGAGGGTGTGGCCAAGTG-----	7032
1402	Qy	uGlyProLeuLysLeuLysPro-----AlaHisGluGlyLeuValAlaThrValLysG1	1420
7033	Db	-----CCACCCAGCTAAGTCCCCCAGCAGGCCCATCACTTCATCCCACTCAGGACACTC	7088
1420	Qy	uAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuPr	1440
7089	Db	GCCATCGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCAACAGCAGCAACAGCGCGC	7148
1440	Qy	oLeuAlaProArgProLeu-----LysGluGlySerIleThrGlnGlyThrProLeuLys	1458
7149	Db	TGCAGCAACAACAATTTGTTGTGTGAAGCGGTTCGATCATCATGGAAACACCG-----	7203
1458	Qy	sTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuIleG1	1478
7204	Db	-----GCCAATAGTCCCCCAGCAACAGATTATCGTTTCATGCACCAGCGCAGCAGCAGC	7256
1478	Qy	ySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAl	1498
7257	Db	AGCACCTCTCTCCCTGTTTCACTAGTCCCAAGTTCGAT-----GG	7292
1498	Qy	aLeuGluArgAlaCysTyrGluGluSerLeuLysSer---ArgProGlyThrAlaSerSe	1517
7293	Db	CTTGTCAGGCAAAACGACACACAGAGGGTGTGGTTCGTGGACCCCGGAGGAGCAAGTGG	7352
1517	Qy	rSerGly-----GlySerIleAlaArgGlyAlaProValIleValPro-----	1531
7353	Db	CAGTGGCAAGCATTGGCTCTATAACGAAGGCACACCGTTGCACATGCGCGCGCATCATTT	7412
1531	Qy	-----	1531
7413	Db	GGAGAGCAACGGCCATACGAGAGCTACTACAAGAGCTCGCAGCGTCACAGTCCCGCCCA	7472
1532	Qy	-----GluLeuGlyLysProArgGln-----SerProLeuThrTy	1543
7473	Db	ACAGCCGGGTGGCAATCAGCAATTTGCCCGCGCGCGCGCAATTCGTTCTCCACAGCGGCC	7532
1543	Qy	rGluAspHisGly-----AlaProPheAlaG1	1552
7533	Db	ACCTCCGAGGGCTATGGTGTGCTGGTGTGAGCTCACCCCTATGCCCGATCGCGGTTCGCCGG	7592
1552	Qy	y-----	1552
7593	Db	TGTGGTGGAAACAGCCGAGGTGCTGACGACGCGCAAAATTTGTGTGTCAGCTACATAAC	7652
1553	Qy	-----Hi-----	1553
7653	Db	CTCACAGCATGTCAGGGACAGCAACAGCAGCAACAGCAGCAGCAGCAGCAACAGCGGCA	7712

Qy	1553	sLeuProArgGlySerProValThr	-----MetArgG1	1564
Db	7713	CATCGCGGCGACAGTGCAGTCGGGTGGAGGTGGCGTGGTTCCGACACGAGATC	7772	
Qy	1564	uProThrProArgLeuGlnGluGlySerLeuSer	-----	1575
Db	7773	ACCATCGCACAAACAGTGTGGCAGCGCGCATGGCTTCGCCATATGGCGGCGACAGGA	7832	
Qy	1576	-----SerSerLysAlaSerGlnAspArgLysLeuTh	1586	
Db	7833	ATCGGCACCGGTGGCGACCGGAGTACTCCAGTCGTGCATCGCACCGCATCATGTCAA	7892	
Qy	1586	rSerThrProArgGluLeuAlaLysSerProHisSerThrValPro	-----	1601
Db	7893	TAGCACATCCG-----AGTCCGCATCGCACCGCGCGCCACAGCGCCCAAGG	7937	
Qy	1602	-----GluHisHis-----ProHisPro	-----	1607
Db	7938	CGTCATCCAGAGGCACAAACCGGCTCCAAAGCCACCATCACCGCTCTCGGCCACCGAG	7997	
Qy	1608	-----IleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTy	1624	
Db	7998	CCGGATGCACATCGCGCGGTATCAATAT-----GCGCCAAGTGTCCAGATGCATT	8048	
Qy	1624	rArgSerHisIleProLeuAlaPheAsp	-----	1633
Db	8049	GGCCTCGTTGTGACGTGGCGGTGCAGACGCCAGTACCGGTACCGTTCGCMAAAGGA	8108	
Qy	1634	-----ProThrSerIleProArgGlyIleProLeuAspAlaAlaAl	1648	
Db	8109	TGACAAAGTCGCGGTCCATCAACGGCGCGGACAAAGTCCCGGATCTGTCACCA--	8166	
Qy	1648	aTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyr	-----	1665
Db	8167	-----CTTGACCATCACCTCTGCCGCCACACCTGTGTGTCGGCGT	8207	
Qy	1666	-----ProPro-----TyrLeuIleLeuArgGlyTyrProAspThrAlaAlaLe	1679	
Db	8208	GGCGCACCACCGCGCGGACCGCACACCGATCAACGGTATCGCGATCTGACCCCTGCA	8267	
Qy	1679	uGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMethHisAsnTh	1699	
Db	8268	TCACCATCATCATACACTGTGTGCAGCAGCAGATT--GCGCAACAACAGCATTATCGTAG	8324	
Qy	1699	rAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSe	1719	
Db	8325	CCTCAATGTGGCGGCCAG--GTCGNATATGCAGCGGCAGATGGATCAGCGGAACGT--	8379	
Qy	1719	rLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHi	1739	
Db	8379	-----	8379	
Qy	1739	sLeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAl	1759	
Db	8380	-----GTATGCG--CCACAGCAGCACCGAGTCCAGCAGACGA	8422	
Qy	1759	aTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerPr	1779	
Db	8423	AGCAACAGCAGCAGCACCAACA	8452	
Qy	1779	oGly-GlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgA	1799	
Db	8453	TGGAACCGGACAGGAGATCGAAGACGGATCGCGACCGGATCTGTACCGGGAAGGG	8512	
Qy	1799	spArgAspArgGlu---ArgAspArgAspArgGluArgGluLys-SerIleLeuThrSer	1817	
Db	8513	AAAGGAGCGTGACGACCGGGAACGAACTGAACGAGAACCGGAACGTGACGGGAGC	8572	
Qy	1818	ThrThrThrValGluHis-----AlaProIleTyrArgProGlyThrGluGlnSerSer	1835	
Db	8573	GCGAGCACGTGAACAGGATCTGTGCGCGCGGTGTGTGTCGCG	8614	

```
QY 1836 GlySerSerGlySerGlyGlyGlyGlySerSerSerArgProAlaSerHisSer 1855
Db 8615 ---AAGAGCGGAGACGATAGTCGACGATGCGGCGCATGTCGCGGCAATGTGTCA 8671
QY 1856 HisAlaHisGlnHisSerProAlaSerProAlaSerGlnAlaLeuGlnArgPro 1875
Db 8672 ---CTGGATCAGGTGGCGCGCGCGGAGAC-----CGTCGCCG 8710
QY 1876 -----SerValLeuHisAsnThrGlyMetLysGlyIle 1886
Db 8711 GTCAATTTCTCGGGCTTCGGTTCGGAACTGCTCCACCACGCTCAA----- 8758
QY 1887 IleThrAlaValGluProSerLysProThrValLeuArgSerThrSer-----Ser 1905
Db 8759 ---TCCAGATCGCGAAGCGGAATCGTACTATCGTCAGGCACACGCGGTCCAGTCCAG 8815
QY 1906 ProValArgProAlaAlaThrPheProAlaThrHisCysProLeuGlyGlyThrLeu 1925
Db 8816 AGNATCGCGCGGCACTCAGCGCCAGAGTCTGATCGATGCCATTA---TCAAGCATG 8872
QY 1926 AspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgVal 1945
Db 8873 AGATCAATCGCTTCCA-----ATGATGCCACCGGTGC-CCGGGTGC----- 8913
QY 1946 AlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProAla 1965
Db 8914 ---GAAATTCGCGGCCCATCGTTCGTCATCTCGGTG-----CGCCACGT 8958
QY 1966 ArgSer-----GlyLeuGluProAlaSerSerProSerLysGlySerGluPro 1981
Db 8959 GGTTCGGATCGGAGCGGCGCACCGGTACAGTCTCTCACCGGCAATGTCTCCATCCG 9018
QY 1982 -----ArgProLeuValProValSerGlyHisAlaThrIleAlaArgThrPro 1998
Db 9019 ATGTACTCGCGATCTACGCGACCGCTCGATGCGGCGGCGGATCAATGCTTACGGC 9078
QY 1999 AlalysAsnLeuAlaProHisAlaSerProAspProAla----- 2013
Db 9079 GAGAACATGCAAGCCC-----AGCTCTCGGGATCGCCAGTGTGATTAATATGAT 9132
QY 2014 -----ProProAlaSerAlaSerAspProHisArgGluLysThrGln 2027
Db 9133 TTGGATCAGGAGCGCATCTCAGCGGCGCGCAGTGTGCCAACAGCAGCAGCAGCAG 9192
QY 2028 SerLysProPheSerIleGlnGluLeuArgSerLeuGlyTyrHisGlySerSer 2047
Db 9193 CAAGCTCCGCTCGTCTCGCAATCTGCGAGCTAGATCCGCG-----CACGTCACCTA 9246
QY 2048 TyrSerProGluGlyValGluProValSerProValSerProSerLeuThrHisAsp 2067
Db 9247 AGGACACCGACTCG---CAATCTGGCGGATCAGCGCCAGTCCACAGCAGATTCATACG 9303
QY 2068 LysGlyLeuProLysHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArg 2087
Db 9304 AAGACATTAACC-----TTTGGGAACTAACCGATTCGATAATTAACAGTGACTACGGC 9357
QY 2088 ProLysGlnProGlyProValLysLeuGlyGluAlaAlaHisLeuProHisLeuArg 2107
Db 9358 ACCAAT----- 2118
QY 2108 Pro-----LeuProGluSerGlnProSerSerPro----- 2118
Db 9376 CCACCATATATGCGCTATTTCAGGAGAGCAGTCAATCTCGCGCGCGATCGCTGGAA 9435
QY 2119 -----LeuLeuGlnThrAlaProGlyValLysGlyHis-----Gln 2130
Db 9436 CAGAACCGTGTATATCAGAGCGCGGAGGAGCGAAGCGAAGCCATTCACAGCAGCAG 9495
QY 2131 ArgValValThrLeuAlaGlnHisIleSerGlnValIleThrGlnAspTyrThrArgHis 2150
Db 9496 CAGCAGCAACACCAGCAGCAACATCACGCCCGCAGCAGCAGCAGCAGCAGCAACAC 9555
QY 2151 HisProGlnGlnLeuSerAlaProLeuProAlaPro-----LeuTyrSerPheProGly--- 2168
Db 9556 CATGCCCAACATCATCCGACAGATGCGGCGCACTGGCTCTGGCTCCAGGTGGA 9615
QY 2169 -----Ala 2169
Db 9616 GCTGTCTCAGGCGGTGGATCGGCGCGCGCGGTAGCGGTGTGTGTGGCGCGCAGACCC 9675
QY 2170 SerCysProValLeuAsp-----LeuArgArgPro----- 2179
Db 9676 AGCACACCTGGGAGATGGGCGCAACATAATCCGAATGCCACAGGCGGTAGTCCGCGA 9735
QY 2179 ----- 2179
Db 9736 AAATTTAATCATGATGATGTACATCATGTGATGGAAACGACGCGGAGCGGTGTGAA 9795
QY 2180 -----ProSerAspLeuTyrLeuProProPro----- 2188
Db 9796 GCTGTCTCAGTCTCTCTACCCAGCGGTGTCTGTTCGCGGAGCAGAGAGAACGCCACG 9855
QY 2189 ---AspHisGlyValProAlaArgGlySerProHisSerGluGlyGlyLysArgSerPro 2207
Db 9856 GGTGTGGCGGAGCGCTCGAGCCGAGTCTCGATCGGAGCGGTGTCTACCACTT 9915
QY 2208 GluProAsn---LysThrSerVal-----LeuGlyGlyGlyGlu 2219
Db 9916 GAAAAATATGTAAGACCGCATCGTGAAGTGTGCGCGATGATATTGGTTATGGAAG 9975
QY 2220 AspGlyLeuGluProValSerProGluGlyMetThrGluPro-----GlyHisSer 2237
Db 9976 AATCGCATGTGCGAGTGGCAACAGAGGATGAATACCGCCGATATGTCGCACATTCG 10035
QY 2238 ArgSerAlaValTyrProLeu-LeuTyrArgAspGlyGluGlnThrGluProSerArgMe 2257
Db 10036 CATGCCGCGTCCATGCTCAGTGTGTGCGGCGC-----ACGACGCC----- 10075
QY 2257 tGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGln 2277
Db 10076 -----ATGTCGCCCATGCGGTGCTATGGAATTGCGAGCAGCA 10113
QY 2277 uSerAsnSerAlaMetValLysSerLysGlnGluLeuAsnLysLysLeuAsnThrHis 2297
Db 10114 AGCAAGAAC-----CA 10125
QY 2297 sAsnArgAsnGluProGluTyrAsnLysSerGlnProGlyThrGluIlePheAsnMetPr 2317
Db 10126 CCGCCGCGAGATCAGTGTGTACGTAAGACGCCCAACCAATACAGAGTGTGTAGATGCC 10185
QY 2317 o-----AlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGln 2334
Db 10186 AGTGTGCGGCGCTCA-GCTGGCAGTGTTCGTTCCGTTTCGTTTCGCGGCCCAATAG 10244
QY 2334 uHisAlaSerThrAsnMetGlyLeuGluAlaIleAlaArgLysAlaLeuMetGlyLysTy 2354
Db 10245 CCACCATTCG----- 10254
QY 2354 rAspGlnTrpGluGluSerProProLeuSerAla----- 2365
Db 10255 -----CCGTATCATCCACCGCGCGGTATGCCCCCAGCAGCCTATCCCTTCCC 10304
QY 2366 ---AsnAlaPheAsnProLeuAsnAlaSerLeuProAlaAlaMetProIle-- 2383
Db 10305 CTACAGCGCCCTGATGTGCGCGGTGCGCGGTGATTCGCCACCGCAGCGGTGCA 10364
QY 2384 -----ThrAlaAlaAspGlyArgSerAspHisTh 2393
Db 10365 GCTAGCCACAGCGGTGGCACCATCGTGTGCTTTTCCCAAGGCCAAGCAGCCCATGC 10424
QY 2393 rLeuThrSerProGly-----GlyGlyGlyLysAlaLysValSerGlyArgProSerSe 2411
Db 10425 CCTGAGTAACCTGGTGCAGTGTGTGGGTGTCTATGTTGGTGGCGCGGCTCTGG 10484
QY 2411 rArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSe 2431
```


Db 10485 A-----GGAAATGCGAGCGGA-----CCAGGTGTGTCTC 10514
Qy 2431 rSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGI 2451
Db 10515 A----- 10515
Qy 2451 uAspArgProSerSerAlaGlySerThrProPheProTyAsnProLeuLeuMetArgLe 2471
Db 10515 ----- 10515
Qy 2471 uGlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLe 2491
Db 10516 -----GTCGGTGTCT-----GGGTATCCGGCGCGCGGACCAGG 10550
Qy 2491 uAla-----GlyProHisHis----- 2496
Db 10551 AAGCGGTGGCGCGCGGTGGCGGTGCACAAATTCAGTTCGCAAGCATCTGTCGGGT 10610
Qy 2497 -----AlaTrpAspGluGluProGlyProLeuLeuLeuCysSerGlnTrpGI 2511
Db 10611 TGGCGTGTCCGCTTTCAGCGCGCGCGCGAGTGCAGTCCAGCGGTGTGTGTCCAGTACGA 10670
Qy 2511 uThrLeuSerAspSerGlu 2517
Db 10671 TGCACCTCAGCGCAGCAAGAT 10689

RESULT 37
LOCUS BD221548 872 bp DNA linear PAT 17-JUL-2003
DEFINITION Human gene and gene expression product v.
ACCESSION BD221548
VERSION BD221548.1 GI:33031318
KEYWORDS JP 2002534055-A/2661.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 872)
Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D., Klinger,J.S.,
Reinhard,C., Giese,K., Randazzo,F., Kennedy,G.C., Pot,D.,
Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M.,
Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V., Jones,L.W.
and Crain,B.S.
Human gene and gene expression product v
Patent: JP 2002534055-A 2661 15-OCT-2002;
CHIRON CORP.HYSEQ INC
OS Homo sapiens (human)
PN JP 2002534055-A/2661
PD 15-OCT-2002
PF 13-MAY-1999 JP 2000548466
PR 14-MAY-1998 US 60/085426,15-MAY-1998 US 60/085537 PR
15-MAY-1998 US 60/085696,21-OCT-1998 US 60/105234 PR
27-OCT-1998 US 60/105877
PI LOUIS T WILLIAMS,JAIME ESCOBEDO,MICHAEL A INNIS,PABLO PI
DOMINGUEZ GARCIA,
PI JULIE SUDDUTH KLINGER,CHRISTOPH REINHARD,KLAUSE GIESE,FILIPPO
PI RANDAZZO,
PI GIULIA C KENNEDY,DAVID POT,ALTAF KASSAM,GEORGE LAWSON,RADOJE
PI DRMANAC,
PI RADONIR CRKVENJAKOV,MARK DICKSON,SNEZANA DRMANAC,IVAN LABAT,
PI DENA LESHKOWITZ,DAVID KITA,VERONICA GARCIA,LEE WILLIAM JONES,
PI BIRGIT STACHE CRAIN
PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC
.C12N5/00,C12O1/68,
PC C12N15/00,C12N5/00
CC Human gene and gene expression product v
FH Key Location/Qualifiers
FT source 1..872
FT /organism='Homo sapiens (human)'.
FEATURES
source 1..872
/organism='Homo sapiens'
/mol_type='genomic DNA'

/db_xref="taxon:9606"

ORIGIN

Alignment Scores:
Pred. No.: 1..12e-14 Length: 872
Score: 1076.00 Matches: 216
Percent Similarity: 82.62% Conservative: 36
Best Local Similarity: 70.82% Mismatches: 34
Query Match: 8.14% Indels: 20
DB: 6 Gaps: 3

US-09-522-753-5 (1-2517) x BD221548 (1-872)

Qy 241 AlaHisArgIleLeuGluGlyProGlnValGluLeuProLeuTyAsnGlnPro 260
Db 10 GCTGTGCAAAATTTTGAAGGTCTTGGCCCAAAAGTTGCACTGCCACTGTATAACGACCA 69
Qy 261 SerAspThrArgGlnTrpHisGluAsnIleGluLeuAlaMetArgLysLeu 280
Db 70 TCAGATACCAAGGTGTACCATGAGAACATCAAGACCAACAGGTGTATGAGGAAAAAACTC 129
Qy 281 IleLeuTyrrPheLysArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGln 300
Db 130 ATTTTATTTTAAAGAGAAATCATGCAAGAAAAACAAAGGGAACAAAAAATCTGCCAG 189
Qy 301 ArgTyrrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnPro 320
Db 190 CGTTATGATCAGCTCTGGAGGATCGGAGAAAAAGTGGACAGAAATAGAAAAATATCTCT 249
Qy 321 ArgArgAlaLysGluSerLysValArgGluTrpTyr-GluLysGlnPheProGluIl 340
Db 250 CGGAGGAAAGCTAAAGAAACAAAACWAGGGAATACTATTTRAAAAAGCAGTTTCCAGAAAT 309
Qy 340 eArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValClyGlnArgGlySerGI 360
Db 310 TCGAAACAAAGAGACAGCAGAAAGATTTCAG---CGAGTTGGGCGAGAGGGAGCTGG 366
Qy 360 YLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSe 380
Db 367 TCITTCAGCCACCATTGCTAGGAGTGAGCATGAGATTCTTCTGAAATATTATTCATGGCTCTC 426
Qy 380 rGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProMetLeuTy 400
Db 427 TGAGCAGGAGATAAATAGAAACAAATTCGCGCAGCTCTC-GTGATTCCACCTATGATGTT 485
Qy 400 rAspAlaAspGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMe 420
Db 486 TGATGAGAACAAAGACGATCAAGTTCATTATCATGAATGGGCTTATGGAGGACCTAT 545
Qy 420 tLysValTyrrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPh 440
Db 546 GAAAGTGATAAAGATAGGACGATTATGATGTTTGGACTGACCATGACCAAGAGAGATCTT 605
Qy 440 eArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluAr 460
Db 606 TAAGGACAAGTTTATTCAGCATCCAAAAAATTTGGCACTAATTCGATCATCATCTGGAGAG 665
Qy 460 gLysThrValAlaGluCysValLeuTyrrTyrrTyrrTyrrTyrrTyrrTyrrTyrrTyrr 480
Db 666 GAAGAGTGTCTCGATTGTTTGTATTACTATTATTAACCAAGAAAAAATGAGAAATTATAA 725
Qy 480 sSerLeuValArgArgSerTyrr---ArgArgArgGlyLysSerGlnGlnGlnGlnGln 499
Db 726 AGCCTCGTCAGAGGAATATTGGGAAACGACAGGAGGAGAACACGACCA----- 774
Qy 499 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 519
Db 775 -----ATTGCTCGACCTCGCAAGAAGA 797
Qy 519 uLysAspGluLysGluLysGluAlaGluLysGluLysGluLysGluLysGluLysGluValGI 539
Db 798 AAAAGTAGAAGAAAAAGAGAGGATTAAGCAGAGAAAAACAAAAAAGAGAGAGAAAA 857

[illegible]

```

Db      63  CCNAGAACTTTGGCCTGATCGCATCATCTTCCTGAGAGGAGAGACAGTCAGTGGCTGCATG 122
Qy      468  LeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyr 487
Db      123  CTCATTACTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 182
Qy      488  ArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 507
Db      183  CGGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 242
Qy      508  GlnGlnGlnProMetProArgSerSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 527
Db      243  CAGCAGCAGCCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 302
Qy      528  GluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAspLeuLeuLys 547
Db      303  GAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 362
Qy      548  GluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSerLys 567
Db      363  GAGAAGACAGACACACTCAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 422
Qy      568  GlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgLysGlyArgLysGlyArgLys 587
Db      423  GCGCGCAAACTGCTCAACAGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 482
Qy      588  AsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMet 607
Db      483  AATGAGGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 542
Qy      608  GluLeuAsnGlu 611
Db      543  GAGCTGAATGAG 554

```

```

RESULT 40
HSU80761/c
LOCUS      560 bp      mRNA      linear      PRI 18-DEC-1997
DEFINITION Homo sapiens CTG26 alternate open reading frame mRNA, complete cds.
ACCESSION  U80761
VERSION     U80761.1  GI:2565090
KEYWORDS   Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 560)
AUTHORS     Margolis,R.L., Abraham,M.R., Gatchell,S.B., Li,S.H., Kidwai,A.S.,
            Breschel,T.S., Stine,O.C., Callahan,C., McInnis,M.G. and Ross,C.A.
            cDNAs with long CAG trinucleotide repeats from human brain
            Hum. Genet. 100 (1), 114-122 (1997)
TITLE       97369492
JOURNAL     PUBMED
MEDLINE     9225980
REFERENCE   2 (bases 1 to 560)
AUTHORS     Margolis,R.L., Abraham,M.R., Gatchell,S.B., Li,S.H., Kidwai,A.S.,
            Breschel,T.S., Stine,O.C., Callahan,C., McInnis,M.G. and Ross,C.A.
            Direct Submission
            Submitted (02-DEC-1996) Psychiatry, Johns Hopkins Univ. Sch. of
            Med., 600 N. Wolfe Street, Meyer 2-181, Baltimore, MD 21205, USA
FEATURES    Location/Qualifiers
             1..560
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /tissue_type="brain"
                /dev_stage="fetus"
                <1..560
                /gene="CTG26"
                <1..375
                /gene="CTG26"
                /note="cysteine rich"
                /codon_start=1
                /product="CTG26 alternate open reading frame"
                /protein_id="AA891452.1"

```

```

/db_xref="GI:2565091"
/translation="SFSSMEASSALCMGVMASLLASLAIERVMRPLRLPWLAVLRP
LEATAPFSSLSSESVSFLRRSSLSFSTSGFSSFSASFSFSSFWLLRGMGC
CCCCCCCCCCCCCWLPRR"
303..341
/repeat_region
/rept_type=tandem
/rept_unit="CTG"

Alignment Scores:
Pred. No.:      6 4e-12      Length:      560
Score:          932.00      Matches:    183
Percent Similarity: 99.46%      Conservative: 0
Best Local Similarity: 99.46%      Mismatches: 1
Query Match:     7.05%      Indels:    0
DB:              9          Gaps:      0

US-09-522-753-5 (1-2517) x HSU80761 (1-560)
Qy      427  GlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMetGln 446
Db      554  CAGGTCATGAACATGTGGAGTGAGCAGAGAGAGAGACCTTCGGGAGAGAGATTCATGCAG 495
Qy      447  HisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCys 466
Db      494  CATCCCAAGAACTTTGGCCTGATCGCATCATTCCTCGAGAGAGAGACAGTGGCTGAGTGC 435
Qy      467  ValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgSer 486
Db      434  GTCTCTATTACTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 375
Qy      487  TyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 506
Db      374  TATCGGCGCGCGGCAAGAGCCAGCAGCAACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 315
Qy      507  GlnGlnGlnProMetProArgSerSerGlnGluLysAspGluLysGluLysGluLysGlu 526
Db      314  CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 255
Qy      527  LysGluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAspLeuLeu 546
Db      254  AAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 195
Qy      547  LysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluLysGluAlaValAlaSer 566
Db      194  AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 135
Qy      567  LysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgLysGlyArgLysThrArgSerMet 586
Db      134  AAAGGCGCAAACTGCTCAACAGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 75
Qy      587  AlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSer 606
Db      74  GCTAATGAGGCAACAGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15
Qy      607  MetGluLeuAsn 610
Db      14  ATGAGCTGAAT 3

RESULT 41
AX753058/c
LOCUS      718 bp      DNA      linear      PAT 23-JUN-2003
DEFINITION Sequence 171 from Patent EPI310567.
ACCESSION  AX753058
VERSION     AX753058.1  GI:32165818
KEYWORDS   Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
REFERENCE   1
AUTHORS     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Stuhlmueller,B. and Haeupl,T.
            Nucleic acid array

```



```
Db 235 TTTGGACATAATTGCATCATCTACTGGAGAGGAGTGTCTCTGATTTGTTTGTATTAC 176
Qy 471 TyrluThrLysLysAsnGluAsnTyrlLysSerLeuValArgArgSerTyr---ArgArg 489
Db 175 TATTTAACCAAGAAATAGATTAATAAGCCCTCGTCAGAGGAATTTATGGGAACGC 116
Qy 490 ArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 509
Db 115 AGAGGCGAAGAACCCAGCA----- 98
Qy 510 GlnProMetProArgSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 529
Db 97 -----ATTGCTGCACCTCGCAAGAAAGAAAGTAAAGAAAGAGGATTAAGCA 44
Qy 530 GluLysGluGluGluLysProGluValGluAsnAspLysGlu 543
Db 43 GAAAAACAGAAAAAAGAGAGAAAGAAAGAGATGAAGAG 2

RESULT 43
BD101314/c
LOCUS BD101314 710 bp DNA linear PAT 27-AUG-2002
DEFINITION Novel genes cloned in humanneuroblastoma and fragments thereof.
ACCESSION BD101314
VERSION BD101314.1 GI:22646888
KEYWORDS WO 0166719-A/3614.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1. (bases 1 to 710)
JOURNAL Nakagawara, A.
COMMENT Novel genes cloned in humanneuroblastoma and fragments thereof
PATENT: WO 0166719-A 3614 13-SEP-2001;
CHIBA PREF. HISAMITSU PHARMACEUTICAL CO INC, AKIRA NAKAGAWARA
OS Homo sapiens (human)
PN WO 0166719-A/3614
PD 13-SEP-2001
PF 02-MAR-2001 WO 2001JP001629
PR 07-MAR-2000 JP 00P 159195
PI AKIRA NAKAGAWARA
PC CL2N15/11, CL12Q1/68, G01N33/53, G01N33/566
CC Novel genes cloned in humanneuroblastoma and fragments thereof
FH Key Location/Qualifiers
FT source 1..710
FT /organism="Homo sapiens (human)".

FEATURES
source
1..710
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 4.46e-10 Length: 710
Score: 845.00 Matches: 174
Percent Similarity: 82.28% Conservative: 35
Best Local Similarity: 68.50% Mismatches: 27
Query Match: 6.39% Indels: 20
DB: 6 Gaps: 3

US-09-522-753-5 (1-2517) x BD101314 (1-710)

Qy 291 ArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGlu 310
Db 710 AGAAACAAAGGNCNCAAAATCTGCACGGTTATGTCAGGCTCATGAGGATG-GAA 652
Qy 311 LysLysValGluArgLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGlu 330
Db 651 AAAAAAGTGCAGATAGAAATAT-CCTCGNAGGAAAGCTAAAGAAAGCAACCAAGG 593
Qy 331 GluTyrTrpLysGlnPheProGluLysGlnArgGluLeuGlnGluArgMet 350
Db 592 GAATACTATGAAAGCGAGTTTCCAGAAATTCGAAACCAAGAGAGACCAAGCAAGAGATTT 533
```

```
Qy 351 GlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHis 370
Db 532 CAG---CGAGTTGGCAGAGCGGAGCTGCTTTTCAGCCACCATCTGCTAGGAGTGAGCAT 476
Qy 371 GluValSerGluLysLysAspGlyLeuSerGluGlnGlnGlnGlnGlnGlnGlnGln 390
Db 475 GRATTTCTCAAAATTTATGATGGCTCTCTGAGCAGGAGAAATATAGAAACAATGCGG 416
Qy 391 GlnLeuAlaValLysProMetLeuTyrAspAlaAspGlnGlnArgGlyLysPheLys 410
Db 415 CAGCTCTCTGATTTCCACCTATGATGTTGATGCGAGAACAAAGACGAGTCAAGTTTCA 356
Qy 411 AsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsn 430
Db 355 AACATGAATGGCTTATGATGGAGACCCCTATGAAAGTGATATAAGATAGGCGAGTTTAT 296
Qy 431 MetTrpSerGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 450
Db 295 GTTTGGACTACCATGAAAGAGAGATCTTTAAGGACAAAGTTTATCCAGCATCCAAAAC 236
Qy 451 PheGlyLeuLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 470
Db 235 TTTGGACTAATTCATCATCTACTTGGAGAGAGAGTGTCTGATGTTGTTGTTGTTATAC 176
Qy 471 TyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyr---ArgArg 489
Db 175 TATTTAACCAAGAAATAGATTAATAAGCCCTCGTCAGAGGAATTTATGGGAACGC 116
Qy 490 ArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 509
Db 115 AGAGGCGAAGAACCCAGCA----- 98
Qy 510 GlnProMetProArgSerGlnGluLysAspGluLysGluLysGluLysGluLysGluAla 529
Db 97 -----ATTGCTGCACCTCGCAAGAGAAAGAAAGTAAAGAAAGAGGATTAAGCA 44
Qy 530 GluLysGluGluGluLysProGluValGluAsnAspLysGlu 543
Db 43 GAAAAACAGAAAAAAGAGAGAAAGAAAGAGATGAAGAG 2

RESULT 44
AX677865 LOCUS AX677865 673 bp DNA linear PAT 27-MAR-2003
DEFINITION Sequence 643 from Patent WO02086122.
ACCESSION AX677865
VERSION AX677865.1 GI:29335270
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Legrain, P. and Daviet, L.
JOURNAL Protein-protein interactions in adipocytes
PATENT: WO 02086122-A 643 31-OCT-2002;
Hybridomics (PR)
FEATURES
source
1..673
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 1.14e-08 Length: 673
Score: 774.50 Matches: 159
Percent Similarity: 80.09% Conservative: 14
Best Local Similarity: 73.61% Mismatches: 38
Query Match: 5.86% Indels: 5
DB: 6 Gaps: 3

US-09-522-753-5 (1-2517) x AX677865 (1-673)
```


Qy 2017 erAlaSerAspProHisArgGluLysThrGln-----SerLysProPheSerIleGlnGlu 2036
Db 108241 GCCAAGCC-----AGGCAAAACACCCAGCCAGCAAGCCA-----GTCAAGCCAC 108285
Qy 2036 euGluLeuArgSerLeuGlyTyHisGlySerSerTySerProGluGlyValGluProV 2056
Db 108286 CCAGCCAGCAAGCGGCCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 108345
Qy 2056 alSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluG 2076
Db 108346 CCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 108387
Qy 2076 luLeuAsp-----LysSerH 2081
Db 108388 GCCAAGATGGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 108447
Qy 2081 isLeuGluGlyGluLeuArgProLysGlnProGly-----p 2093
Db 108448 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 108507
Qy 2093 roValLysLeuGlyGlyGluAlaAlaHis-----LeuProHisLeuArgP 2108
Db 108508 CAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 108567
Qy 2108 roLeuProGluSerGlnProSer-----SerSerProLeuLeuGlnThrAlaProGlyV 2126
Db 108568 CCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 108627
Qy 2126 alLysGlyHisGlnArgValThrLeuAlaGlnHisIleSerGluValIleThrGlnA 2146
Db 108628 GCCAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 108687
Qy 2146 spTyThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTy----- 2164
Db 108688 GCCACCAAGACACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 108747
Qy 2165 -----SerPheProGlyAlaSerCysProValLeuAspLeuArgPro-ProSer 2181
Db 108748 GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 108780
Qy 2182 AspleuTyLeuProProAspHisGlyAlaProAlaArgGlySerProHisSerGlu 2201
Db 108781 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 108840
Qy 2202 GlyGlyLysArgSerProGluProAsnLysThrSer-----ValLeuGly 2216
Db 108841 CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 108900
Qy 2217 GlyGlyGluAspGlyIleGluProValSerProGluGlyMetThrGluProGlyHis 2236
Db 108901 CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 108960
Qy 2237 SerArgSerAlaValTyProLeuLeuTyArgAspGlyGluGlnThrGluProSerArg 2256
Db 108961 CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 108999
Qy 2257 MetGlySerLysSerProGlyAsnThrSerGlnProAlaPheSerLysLeuThr 2276
Db 109000 CCACACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 109059
Qy 2277 GluSerAsnSerAlaMetValLysSerLysGlnGlnLeuLeuAsnLysLysLeuAsnThr 2296
Db 109060 CAAGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 109119
Qy 2297 HisAsnArgAsnGluProGluTyArgAsnIleSerGlnProGluThrGluIlePheAsnMet 2316
Db 109120 GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 109170
Qy 2317 ProAlaIleThrGlyThrGlyLeuMetThrTyArgSerGlnAlaValGlnGluHisAla 2336
Db 109171 -----AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 109191
Qy 2337 SerThrAsnMetGlyLeuGluAlaIleArgLysAlaLeuMetGlyLysTyAspGln 2356

Db 109192 AAGCCAGGCATGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 109230
Qy 2357 TrpGluGluSerProLeuSer-----AlaAsnAlaPheAsnProLeu 2371
Db 109231 ---CAGCAAAACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCT 109287
Qy 2372 AsnAlaSerAla-----SerLeuProAlaAlaMetProIle 2383
Db 109288 AAGCCAGCGGTAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 109347
Qy 2384 ThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyAla 2403
Db 109348 CGGCCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 109407
Qy 2404 LysValSerGlyArgProSerArgLysAlaLysSerProAlaProGlyLeuAlaSer 2423
Db 109408 AGCTAGCTGGCAGCTGCGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 109467
Qy 2424 GlyAspArgProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThr 2443
Db 109468 GGCC--AGGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 109522
Qy 2444 ProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPhe--- 2462
Db 109523 CAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 109582
Qy 2463 -----ProTyAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSer 2478
Db 109583 AAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 109641
Qy 2479 ProProProProGlyLeuProAla-----GlySerGlyProLeuAlaGlyProHisHis 2496
Db 109642 CCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 109701
Qy 2497 AlaTrpAspGluProLysPro 2504
Db 109702 CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 109725
RESULT 46
AC130376 150769 bp DNA linear HTG 27-MAR-2003
LOCUS Homo sapiens chromosome 15 clone RP13-70104 map 15, 4 unordered
DEFINITION pieces.
AC130376
VERSION AC130376.9 GI:27369452
KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_CANCELLED.
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)
REFERENCE 1 (bases 1 to 150769)
AUTHORS Birren, B., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 15, clone RP13-70104
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 150769)
AUTHORS Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barra, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrum, J., Meneus, L., Minova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, J., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,


```
Qy 834 aproProValGluGluGluGlu 842
Db 922 CCCAGGCGCAGGCGAGGAGGAGCAG 946

RESULT 48
AX396270 LOCUS 520 bp DNA linear PAT 18-MAY-2002
DEFINITION Sequence 485 from Patent WO0212328.
ACCESSION AX396270
VERSION AX396270.1 GI:21067017
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
JOURNAL King G.E., Meagher M.J., Xu, J. and Secrist, H.
COMMENT Compositions and methods for the therapy and diagnosis of colon
cancer
PATENT: WO 0212328-A 485 14-FEB-2002;
CORIXA CORPORATION (US)
FEATURES
source
1..520
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 3.41e-08 Length: 520
Score: 746.50 Matches: 141
Percent Similarity: 91.95% Conservative: 19
Best Local Similarity: 81.03% Mismatches: 13
Query Match: 5.65% Indels: 1
DB: 6 Gaps: 1

US-09-522-753-5 (1-2517) x AX396270 (1-520)

Qy 285 LysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGln 304
Db 1 AAAAAAGAAATCATGCGAAGAAACAAAGGGAACAAAATCTGCCAGCGTTATGATCAG 60

Qy 305 LeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgAla 324
Db 61 CTATGAGGCGATGGGAGAAAAAAGTGCACAGAATAAATAATCTCTGAGGAGAAAGCT 120

Qy 325 LysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluLysGlnArg 344
Db 121 AAAAGAAAGCAAAACAGGGAATATCTATGAAAGCAGTTTCCAGAAATTCGAAAACAAAGA 180

Qy 345 GluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSer 364
Db 181 GAACAGCAAGAAAGATTTCAG--CGAGTTGGGCGAGGGAGGCTGCTTTTCAGCCACC 237

Qy 365 AlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsn 384
Db 238 ATTGCTAGGAGTGAGCATGAGATTCTGAAATATTATTCATGGCTCTCTGAGCAGGAGAT 297

Qy 385 LeuGluLysGlnMetArgGlnLeuAlaValIleProMetLeuTyrAspAlaAspGln 404
Db 298 AATGAGAAACAAATGCGCGAGCTCTCTGTGATTCCACCTATGATGTTGTATGACAGAACAA 357

Qy 405 GlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLys 424
Db 358 AGACGAGTCAAGTTTCAATTAATGAAATGGGCTTATGAGGACCCCTATGAAAGTATATAA 417

Qy 425 AspArgGlnValMetAsnMetTrpSerGluGlnGlnLysGluThrPheArgGluLysPhe 444
Db 418 GATAGGCAGTTTATGATGTTTGGACTGACCATCGAAAGGAGATCTTTTAAGGACCAAGTTT 477

Qy 445 MetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeu 458
Db 478 ATCCAGCATCCAAAAAATTTGGACTTAATTTGCATCATCTTG 519
```

RESULT 49

```
AC020019/c LOCUS 40871 bp DNA linear HTG 03-JAN-2000
```

```
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC020019
VERSION AC020019.1 GI:6664878
```

```
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Drosophila melanogaster (fruit fly)
```

```
ORGANISM Drosophila melanogaster
```

```
KEYWORDS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
```

```
REFERENCE 1 (bases 1 to 40871)
```

```
AUTHORS Adams, M. and Venter, J.C.
```

```
TITLE Direct Submission
```

```
JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
```

```
Rockville, MD, USA
```

```
COMMENT This sequence was identified as CDM:10211762 by the submitter.
```

```
For more information on this record e-mail to fly@celera.com.
```

```
* NOTE: This is a 'working draft' sequence.
```

```
* This sequence will be replaced
```

```
* by the finished sequence as soon as it is available and
```

```
* the accession number will be preserved.
```

```
FEATURES
```

```
source
```

```
1..40871
```

```
/organism="Drosophila melanogaster"
```

```
/mol_type="genomic DNA"
```

```
/db_xref="taxon:7227"
```

```
ORIGIN
```

```
Alignment Scores:
```

```
Pred. No.: 1.34e-06 Length: 40871
```

```
Score: 743.00 Matches: 237
```

```
Percent Similarity: 41.08% Conservative: 106
```

```
Best Local Similarity: 28.38% Mismatches: 254
```

```
Query Match: 5.62% Indels: 238
```

```
DB: 2 Gaps: 23
```

```
US-09-522-753-5 (1-2517) x AC020019 (1-40871)
```

```
Qy 144 LysLeuGluProValSerProSerProSerProHisThrAspProGluLeuGluLeuVal 163
```

```
Db 11990 CAGGTGGAGGCCATTTTCGCCGACATTCGCC-----AGCCATAGTTCGATTGAGGCGC 11937
```

```
Qy 164 ProProArgLeuSerLysGluLeuLeuGlnAsnMetAspArgValAspArgGluLeu 183
```

```
Db 11936 GGAAGGACCGCCGCAAGAGGATCTCTCATCAATCCAAAGGTTGACATGAGATC 11877
```

```
Qy 184 ThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeuGluGluGlu 203
```

```
Db 11876 AATCCGCTGAGACGACATGGAACGTTGCGCAAAAGGAGAAATCCCTCATGAGGAG 11817
```

```
Qy 204 Ala-----AlaLysProGluProGluProGluProValSer----- 215
```

```
Db 11816 GCGCGCTGCGCAAGAGCAGAGCGCGCCCAAGAGTTGAACGATCAACAACATGATCAG 11757
```

```
Qy 216 ProProIleGluSerLysHisArgSer-----LeuValGlnIleIleTyrAspGlu 233
```

```
Db 11756 GAACCATGATGTGAACATCTCTGCGCGACGCAAAATGCTACGCGAGAGATCTATGCGGCC 11697
```

```
Qy 234 AsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluLysLeuGlyProGlnValGlu 253
```

```
Db 11696 AATCGGAAGACGCGCCCAAGGCAACATTCATCTGCTGAGATGATCAGCGCGCGATCG 11637
```

```
Qy 254 -----LeuProLeuTyrAsnGlnProSerAspThr 263
```

```
Db 11636 TCTCCGGGTTCCGTTGCGCGACAGACCATGTTGCCATTGTATTAACACGACCTCGATGTC 11577
```

```
Qy 264 ArgGlnTyrHisGluAsnIleLysIleAsnGlnAla---MetArgLysLysLeuIleLeu 282
```

```
Db 11576 GAGGCATCTGCCATGTGTAATACCGACCATCAGATCAAAATTCGTGCGCGCTACTTTTG 11517
```

Qy	283	TyrPheLysArgArgAsnHisalaIargLysGlnTrpLysGlnLysPheCysGlnArgTyr	302
Db	11516	CAATACGAAAAATTGAAGCCCAACGCTGGCGCACCAATCAGGACCTGGTGGAGAGTAC	11457
Qy	303	AspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArg	322
Db	11456	ACCAAGGATCAGCGGATTGGCAGCGCATGTCGAGCGTATGGAGGCCAGTGCACAGAGG	11397
Qy	323	ArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLys	342
Db	11396	AAAGCAGCGGAGCAAAAAATCGCAATTCTTCGAAAAAGTCTTCAACGAGCTAAGGAAA	11337
Qy	343	GluArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSer	362
Db	11336	CAGCGCAGGATGAAGAAGCTTTC--AATCGCTCGGTTCCGCG--	11295
Qy	363	MetSerAlaAlaArgSerGluHisGluValSerGluIleLeuAspGlyLeuSerGluGln	382
Db	11294	-----NTTAAATCCGAGCGGATCTCGAGGAGATCATGGATGGCTCGAGGACGAG	11244
Qy	383	GluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAla	402
Db	11243	CGCTTGAGGATAAAGAAGATCGCATCCTATGCGGTATACCGCTCTGATCACGATGCC	11184
Qy	403	AspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysVal	422
Db	11183	CGCCAGGCGGATGCGGCTATCATATGAATGACTTATCGAGGATATGGTGGCAGTT	11124
Qy	423	TyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGlu	442
Db	11123	CATCAACAGCGTAAGGCCCTCAACATGTGCGACCGCGCGGAGAGAGACCTTTAAGGAG	11064
Qy	443	LysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThr	462
Db	11063	AAGTATTTCAGCATCCGAAAACTTTGAGCCATTGCGCTAGTTTGGATCGCAATCG	11004
Qy	463	ValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeu	482
Db	11003	CCACAGGATTGCGTTGCTACTACTACCTCAGCAAGAACGACGAGAATTACAAGCAGCTG	10944
Qy	483	ValArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGln	502
Db	10943	CTTCGAAATCCCGCAACGACGCGCAGTTCCGCTAATCCAGCCAGGCTCAGGCCCA	10884
Qy	503	Gln-----GlnGlnGln	506
Db	10883	CAGCCGAGTCATTATCGATTTCGATGACCACTGGCGTTTATGACGCGATTGACGCGCA	10824
Qy	507	GlnGlnGlnGlnProMetProArgSerSerGln--GluGluLysAspGluLysGluLys	525
Db	10823	CAGCAACAAAAAGTCCGCGGTCTCATCCGCGTTGCTGAGCGTGGCGTCCGGAACGT	10764
Qy	526	GluLysGluAlaGluLysGluGluLysProGluValGluAsnAspLysGluAspLeu	545
Db	10763	CGACCGAGCTGAGCGTGTGGCAGAGAACGCGAGCTGATCTGCCAAA-----	10713
Qy	546	LeuLysGluLysThrAspAspThrSerGlyLysAsnAspGluLysGluAlaValAla	565
Db	10712	-----GCGGCTGAAAGTCCGCGCAAGGCAAGCCCGCC	10677
Qy	566	SerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThrArgSer	585
Db	10676	ACTAAG-----	10671
Qy	586	MetAlaAsnGluAlaAsnSerGluAlaIleThrProGlnGlnSerAlaGluLeuAla	605
Db	10670	---GCTGTCGAAGCAACGCT--	10653
Qy	606	SerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLys	625
Db	10653	-----	10653
Qy	626	GlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThr	645

Db	10653	-----	10655
Qy	646	ValSerGlnCysIysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIle	665
Db	10653	-----	10653
Qy	666	LeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLys	685
Db	10652	-----GCCGTGTAAGTGGCCAAA	10632
Qy	686	AlaProAlaAlaAlaSerGluGluAlaAlaPheProValValGluAspGluMet	705
Db	10631	CGCGTGCAGCAGCGCGCGCAGCAGCAACTACCGCCACACCGCAACACACACACA	10572
Qy	706	GluAlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHis	725
Db	10571	TCATCATCAACATCATCATCATCATCA	10542
Qy	726	AlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSer	745
Db	10541	GCATCATCAGCATCAACAGCATCATCATCGCGCAAGTCCAGCGACATGTG	10491
Qy	746	SerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsn	765
Db	10490	-----GCAGGAATAGCCAGATAAAGCGATGCCGCGAAA	10443
Qy	766	GlyProLysProAlaThrLeuGlyAlaAspGlyProProGlyProProThrPro	785
Db	10442	AGCATTAAGTAATGCAGCAACAGTGTG	10410
Qy	786	ProArgArgThrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAla	805
Db	10409	-----ACGGCAACGGGAAGT	10395
Qy	806	ProThrProProAlaProProSerProSerAlaProProValProLysGlu	825
Db	10394	CCGACAGCAACTACACTGCAACAGCAACAGCACCA	10353
Qy	826	GluLysGluGluThrAlaAlaAlaProProValGluGluGluGluGluGluLysPro	845
Db	10352	-----GAGATCTCAGCGGGC	10320
Qy	846	ProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluProValLysSerGlu	865
Db	10319	AATGCCAGAGGAGGAGGAGTGCACACGCGGAGCAGCA	10281
Qy	866	CysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAla	881
Db	10280	-----ACGGTGGCACTGCGGTATACACCGCAACCGGAGCAAGTCCGCCAGTCTGCGCAG	10222
Qy	882	AlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGly	901
Db	10223	CGCAACACGGCAACCGGCGCCACGGCCACAGCTGCCGCCAAGGTGTTGGCAACCGGAA	10166
Qy	902	ArgAla	916
Db	10163	ACAGCGACTGAGCCAGCGGGAACGGTGTCTAAGGAGCTGACTCTCGGCAGATGCCAAC	10104
Qy	917	SerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGly	931
Db	10103	GATCCGCTGGCCAAACTGCTTTCGAAGGCTATCAATCCGAGGCG	10059
RESULT	50		
AC023741/c			
LOCUS	160440 bp	DNA	linear
DEFINITION	Drosophila melanogaster X BAC RP98-2309 (Roswell Park Cancer		INV 18-JUN-2002
	Institute Drosophila BAC Library) complete sequence.		
ACCESSION	AC023741		
VERSION	AC023741.4	GI:21450410	
KEYWORDS	HTG.		
SOURCE	Drosophila melanogaster (fruit fly)		
ORGANISM	Drosophila melanogaster		

Db	70921	----	GCTGTCGAAGCAACGGCT	-----	70904
Qy	606	SerMetGluLeuAsnGluSerSerArgTrpThrGluGluMetGluThrAlaLysLys	625		
Db	70904	-----		-----	70904
Qy	626	GlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThr	645		
Db	70904	-----		-----	70904
Qy	646	ValSerGlnCysLysAsnPheThrPheAsnTyrLysArgGlnAsnLeuAspGluIle	665		
Db	70904	-----		-----	70904
Qy	666	LeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLys	685		
Db	70903	-----		-----	70983
Qy	686	AlaProAlaAlaLysSerGluGluAlaAlaPheProValValGluAspGluGluMet	705		
Db	70882	GCGGCTGCAGCAGCGCGCGGCGAGCAACATCTACCGCCACACCGCAACACCA	70823		
Qy	706	GluAlaSerGlyValSerGlyAsnGlnGluGluMetValGluGluAlaGluAlaLeuHis	725		
Db	70822	TCATCATCAACATCATCATCATCATCATCA	-----	70793	
Qy	726	AlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSer	745		
Db	70792	GCATCATCAGCATCAACAGCATCATCATCGCGCAAGTCCAGCACATG	-----	70742	
Qy	746	SerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsn	765		
Db	70741	-----		-----	70694
Qy	766	GlyProLysProProAlaThrLeuGlyAlaAspGlyProProGlyProProThrPro	785		
Db	70693	AGCGATAAGAATCCAGCAACAGCTGCT	-----	70661	
Qy	786	ProArgThrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAla	805		
Db	70660	-----		-----	70646
Qy	806	ProThrProProAlaProProSerProSerAlaProProValValProLysGlu	825		
Db	70645	CCGACAGCAGCACTACCTGCAACAGCAACAGCACCA	-----	70604	
Qy	826	GluLysGluGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysPro	845		
Db	70603	-----		-----	70571
Qy	846	ProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluProValLysSerGlu	865		
Db	70570	AATGCCGAGGAGGAGCGAGCTGCCACAGCGGAGGAGCA	-----	70532	
Qy	866	CysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAla	881		
Db	70531	-----		-----	70475
Qy	882	AlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysGlyGlySerGly	901		
Db	70474	GCACACACCGCAACCGGGCGCCACGGCGACAGCTGCCCAAGGGTGTGGCAAA	70415		
Qy	902	ArgAla	-----	916	
Db	70414	ACAGCACTGAGCCAGCGGAACGGCTGTATAGGGAGCTGACTTCGGCCAGATGCCAAC	70355		
Qy	917	SerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGly	931		
Db	70354	GATCCGCTGCCAAAACCTGCTTCGAGGGCTATCAATGCCGAGGGC	70310		
RESULT 51					
AC104627/c					
LOCUS	AC104627	187921 bp	DNA	linear	INV 22-MAY-2002

RESULT 51
AC104627/
LOCUS

187921 bp DNA linear INV 22-MAY-2002

DEFINITION	<i>Drosophila melanogaster</i> X BAC RP98-4A11 (Roswell Park Cancer Institute <i>Drosophila</i> BAC Library) complete sequence.
ACCESSION	AC104627
VERSION	GI:21070446
KEYWORDS	HTG.
SOURCE	<i>Drosophila melanogaster</i> (fruit fly)
ORGANISM	<i>Drosophila melanogaster</i>
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; <i>Drosophila</i> .
AUTHORS	1 (bases 1 to 187921) Muzny, D., Scherer, S., Adams, M.D., Holt, R.A., Evans, C.A., Gocayne, J.D., Tabor, P., Williamson, A., Homs, F.H., Dugan-Rocha, S.D., Sodergren, E.S., Hodgson, A.H., Chen, R.C., Ayele, M., Scott, G.S., Worley, K.W., Amaratunga, P.G., Brandon, R.C., Rogers, Y., An, H., Baldwin, D., Beeson, K.Y., Brown, M., Buhay, C., Busam, D.A., Center, A.A., Chen, G., Chen, Z., Clerc-Blankenbush, K., Davenport, L.B., Dietz, S.M., Ding, Y., Dodson, K., Doup, L.E., Draper, H., Emery-Cohen, A., Ferreria, S., Garg, N.D.S., Houck, J., Hostin, D., Howland, T.J., Hume, J.J., Ibegwam, C., Jatali, M., Kovar, C., Liu, W., Mattei, B., McIntosh, T.C., Morgan, M., Moy, M., Murphy, B., Nelson, K.A., Ndassa, Y., Nguyen, N., Perez, L., Pittman, G.S., Puri, V., Scheeler, F., Shen, H., Strong, R., Tector, C., Wang, Q., Williams, S.M., Xiang, J., Zaveri, J.S., Zhou, J., Zorrilla, S., Smith, H.O., Wheeler, D., Weinstock, G., Gibbs, R. and Venter, J.C.
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 187921)
AUTHORS	Wortley, K.C., Adams, C., Adio-Oduola, B., Ali-oesman, F.R., Allen, C., Alsbrooks, S.L., Amarutunge, H.C., Are, J.R., Banks, T., Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowte, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J.J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loughheed, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Monabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quites, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Wu, C., Wu, Y., Wu, Y., Zhou, J., Zorrilla, S., Nelson, J.C., Weinstock, G. and Gibbs, R.
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 187921)
AUTHORS	Worley, K.C.
TITLE	Direct Submission
JOURNAL	Submitted (15-DEC-2001) Human Genome Sequencing Center, Department

Db 6448 CCACAGATTCCGTTCCGTACTACTACTACCTCAGCAAGACCGAGAAATTACAAGCAGCTG 6389
 Qy 483 ValArgArgSerTyrArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGln 502
 Db 6388 CTTTGAATAATCCCGCAACGACGCGAGTTTCGGTAAATCCAGCAAGGCTCAGCGCGCA 6329
 Qy 503 Gln-----GlnGlnGln 506
 Db 6328 CAGCCGCGAGTGCATTATCGATTTCGATGACCACTGGCGTTATGACGCGATTTCAGCGCGAA 6269
 Qy 507 GlnGlnGlnGlnProMetProArgSerSerGln---GluGluLysAspGluLysGluLys 525
 Db 6268 CAGCAACAAAGATCCGCGGTCGTTTCATCCGCCCTGCTGAGCGTGCAGCGCGCAACGT 6209
 Qy 526 GluLysGluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAspLeu 545
 Db 6208 GCCGCCGAGCGTGAGCGTGTGGCAGAGAAACGCGCAGCTGATGCTGCCAA----- 6158
 Qy 546 LeuLysGluLysThrAspThrSerGlyGluAspAsnAspGluLysGluAlaValAla 565
 Db 6157 -----GCGGCTGAAAGTCCCGCCGAAAGGCAAGCGCGGCC 6122
 Qy 566 SerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgLysThrArgSer 585
 Db 6121 ACTAAG----- 6116
 Qy 586 MetAlaAsnGluAlaAsnSerGluAlaIleThrProGlnGlnSerAlaGluLeuAla 605
 Db 6115 ---GCTCGAAGCAACGGCT----- 6098
 Qy 606 SerMetGluLeuAsnGluSerSerArgTrpThrGluGluMetGluThrAlaLysLys 625
 Db 6098 ----- 6098
 Qy 626 GlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThr 645
 Db 6098 ----- 6098
 Qy 646 ValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIle 665
 Db 6098 ----- 6098
 Qy 666 LeuGlnGlnHisLysLysLeuLysMetGluLysGluArgAsnAlaArgGlyLysLysLys 685
 Db 6097 -----GCGGTGAAAGAGGTGGCCAAA 6077
 Qy 686 AlaProAlaAlaSerGluAlaAlaPheProProValValGluAspGluLysMet 705
 Db 6076 CGCGCTCGACGAGCGCGCGGCGGACGACCACTACCGCCCAACCCGCAACCAACCA 6017
 Qy 706 GluAlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHis 725
 Db 6016 TCATCATCAATCATCATCATCATCA----- 5987
 Qy 726 AlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnSer 745
 Db 5986 GCATCATCAGCATCAACAGCATCATCATCAGGCGCAAGTCCAGCGCATTG----- 5936
 Qy 746 SerAspThrGluSerIleProSerProHisThrGluAlaLysAspThrGlyGlnAsn 765
 Db 5935 -----CGAGAAATAGCGCGAGATAAAGCATCCCGGCAAA---ACGGCTAGCGCC 5888
 Qy 766 GlyProLysProProAlaThrLeuGlyAlaAspGlyProProGlyProProThrPro 785
 Db 5887 AGCGATAAGATCGAGCAACAGCTGTT-----GGACCA----- 5855
 Qy 786 ProArgArgThrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAla 805
 Db 5854 -----ACGGCAACCGGAACCT 5840
 Qy 806 ProThrProProAlaProProSerProSerAlaProProProValProLysGlu 825
 Db 5839 CCGACAGCAGCAACTACACCTGCAACGCAACCAACGACCA----- 5798

Qy 826 GluLysGluGluThrAlaAlaProProValGluGluGlyGluGluGlnLysPro 845
 Db 5797 -----GAGATCTCAGCGGC-----GGCGAGCCCAAGAGCAA 5765
 Qy 846 ProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGlu 865
 Db 5764 AATGCCGAGGAGGAGGAGTGCACACGCGCGCAGCA----- 5726
 Qy 866 CysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAla-----Glu 881
 Db 5725 ---ACGGTGGCCACTCGCGGTACACCGGCAACCGGAGCAAGTGGCGCAGTGTGGCGGAG 5669
 Qy 882 AlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGly 901
 Db 5668 GCGACACGCAACGGCGGCCACCGCCACACGCTGCCCCCAAGGGTGTGGCAACCGGAA 5609
 Qy 902 ArgAla-----ThrThrAlaLysSerSerGlyAlaProGlnAspSerAsp 916
 Db 5608 ACAGCGACTGAGCGAGCGGGAACGGCTGCTAAGGAGTGACTCTCGGCCAGATGCCAAC 5549
 Qy 917 SerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGly 931
 Db 5548 GATCCGCTGCCAAAACCTGCTTGAAGGCTATCAATGCCGAGGC 5504

RESULT 52

AE003490/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AE003490 302665 bp DNA linear INV 14-FEB-2003
 Drosophila melanogaster chromosome X section 42 of 74 of the
 complete sequence.

AE003490 AE002593 AE014298
 AE003490.2 GI:22832147

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 302665)

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,

Ananides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.P.,

George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,

Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Hendrix, L.X.,

Brannon, R.C., Rogers, Y.H., Blazee, R.G., Champe, M., Pfeiffer, B.D.,

Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L.,

Abail, J.F., Agbayani, A., An, H.J., Andrews-Pfannkuch, C., Baldwin, D.,

Ballouz, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M.,

Beeson, K.Y., Benos, P.V., Betman, B.P., Bhandari, D., Bolshakov, S.,

Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P.,

Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A.,

Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B.,

Davies, P., de Pablo, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I.,

Dietz, S.M., Dodson, K., Dou, P.L.E., Downes, M., Dugan-Rocha, S.,

Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C.,

Ferreira, S., Fleischmann, W., Fowler, C., Gabriellian, A.E., Garg, N.S.,

Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z.,

Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J.,

Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,

Wei, M.H., Ibegwam, C., Jalili, M., Kalush, F., Karpen, G.H., Ke, Z.,

Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,

Kravitz, S., Kulp, D., Lai, Z., Lasoko, P., Lei, Y., Levitsky, A.A.,

Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattel, B., McIntosh, T.C.,

McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C.,

Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,

Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K.,

Nusskern, D.R., Pacle, J.M., Palazzolo, M., Pittman, G.S., Pan, S.,

Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K.,

Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I.,

Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C.,

Stapleton, M., Strong, R., Sun, E., Svirska, R., Tector, C., Turner, R.,

Venter, A., Wang, A.H., Wang, X., Wang, Z.Y., Waeatman, D.A.,

Weinstein, G.M., Weissenbach, J., Williams, S.M., Woodgett,

Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F.,

Yeh, R.F.,

Yeh, R.F.,

Yeh, R.F.,

Yeh, R.F.,

Yeh, R.F.,

Yeh, R.F.,

Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)
20196006
10731132

REFERENCE

AUTHORS

2 (bases 1 to 302665)

Celniker,S.E., Adams,M.D., Krommiller,B., Wan,K.H., Holt,R.A.,
Evans,C.A., Gocayne,J.D., Ananides,P.G., Brandon,R.C., Rogers,Y.,
Banzon,J., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,K.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,I.J.,
Ibegwam,C., Jallali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Paclob,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Tyler,D.,
Williams,S.M., Zaveri,J.S., Smith,H.O., Venter,J.C. and Rubin,G.M.
Sequencing of *Drosophila melanogaster* genome
Unpublished

3 (bases 1 to 302665)

Misra,S., Crosby,M.A., Matthews,B.B., Bayraktaroglu,L.,
Campbell,K., Hradecky,P., Huang,Y., Kaminker,J.S., Prochnik,S.E.,
Smith,C.D., Tupy,J.L., Bergman,C.M., Berman,B.P., Carlson,J.W., de
Celniker,S.E., Clamp,M.E., Drysdale,R.A., Emmert,D., Frise,E., de
Grey,A.D.N.J., Harris,N.L., Krommiller,B., Marshall,B.,
Millburn,G.H., Richter,J., Russo,S., Searle,S.M.J., Smith,E.,
Shu,S., Smucniak,F., Whitfield,E.J., Ashburner,M., Gelbart,W.M.,
Rubin,G.M., Mungall,C.J. and Lewis,S.E.
Annotation of *Drosophila melanogaster* Genome
Unpublished

REFERENCE

AUTHORS

4 (bases 1 to 302665)

Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
5 (bases 1 to 302665)

REFERENCE

AUTHORS

FlyBase

Direct Submission
Submitted (06-SEP-2002) University of California Berkeley, 539 Life
Sciences Addition, Berkeley, CA 94720, USA
6 (bases 1 to 302665)

REFERENCE

AUTHORS

FlyBase

Direct Submission
Submitted (23-JAN-2003) University of California Berkeley, 539 Life
Sciences Addition, Berkeley, CA 94720, USA
On Sep 13, 2002 this sequence version replaced gi:7292788.

COMMENT

FEATURES

Location/Qualifiers

1. 302665
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/chromosome="X"
/note="genotype: Y[1]; cn[1] bw[1] sp[1]; Rh6[1]"
3278..4294
/locus_tag="CG15728"
/note="last curated on Fri Jun 28 07:20:51 PDT 2002"
/map="11B5-11B5"
/db_xref="FLYBASE:FBgn0030409"
join(3278..3409,3466..3657,3722..4294)
/locus_tag="CG15728"
/product="CG15728-RA"
/db_xref="FLYBASE:FBgn0030409"
join(3278..3409,3466..3657,3722..4294)
/locus_tag="CG15728"
/note="CG15728 gene product"
/codon_start=1
/product="CG15728-PA"

/protein_id="AAF48183.1"
/db_xref="GI:7292789"
/db_xref="FLYBASE:FBgn0030409"
/translation="MAEYNMHWPQGSLEELRQTMRSRAIRQCYFIYATGNILE
IISGDELLNQGIQFQADKLAAYVCGIMVLLHHEDTSATQVRTIFLQKCFYMA
CSSESHIQVCHVGLDIDSSSTMLNLILGCRVASPLCTWASVAVLLWMLDHH
SDGLDHLRLPANTLLVAVVKPHIYIYSYLSHLVLISSILLIIPGNAOQQ
LCQETDVPEDYMKLARDCCSILRWLIAIVDELRLPMVNNDLGHLHERLVLESTICE
LWQLLGHVLCYQKSGPOCK"
complement(4415..5643)
/locus_tag="CG15727"
/note="last curated on Fri Jun 28 12:19:58 PDT 2002"
/map="11B5-11B5"
/db_xref="FLYBASE:FBgn0030410"
complement(join(4415..5455,5520..5643))
/locus_tag="CG15727"
/product="CG15727-RA"
/db_xref="FLYBASE:FBgn0030410"
complement(join(4600..5455,5520..5548))
/locus_tag="CG15727"
/note="CG15727 gene product"
/codon_start=1
/product="CG15727-PA"
/protein_id="AAF48184.1"
/db_xref="GI:7292790"
/db_xref="FLYBASE:FBgn0030410"
/translation="MSGKDDRSKERNMKQYNHHKKDQGRSTSSGSPPTSRSRM
VDVAVSSDEPLRARGNQSSGAGRSSLAQNRNMDLTPSDTDDVEQLDGDAPID
ENARQAGDFQQAQPSFLGCGHFTFGSEREWANVAGQTKLTKAASAYFTLNLT
RLNVGLQITPLIKRMDYPSLFRQIAAQAKAERAAEAYVQOCLKDANGAKSRAP
SAKSNKDVAPAAAEKPIAASAEPELDELAMTDTQLDIGSTTITMPPMVQAPATPS
SAASKNGVEWLDSDLDE"
5955..7164
/locus_tag="CG2540"
/note="last curated on Fri Jun 28 07:21:09 PDT 2002"
/map="11B5-11B5"
/db_xref="FLYBASE:FBgn0030411"
5955..7164
/locus_tag="CG2540"
/product="CG2540-RA"
/db_xref="FLYBASE:FBgn0030411"
6004..6939
/locus_tag="CG2540"
/note="CG2540 gene product"
/codon_start=1
/product="CG2540-PA"
/protein_id="AAF48185.1"
/db_xref="GI:7292791"
/db_xref="FLYBASE:FBgn0030411"
/translation="MLDFTLESRSGSRADQLPGNITYTEFFANVIEPREHGCKSADQ
DFTQGSMLRTFTDENSENLEPPAATDADWIFGYSLVKTKDFFYIDRRRGFWGFK
RFPYQSHIDHRIPEPGRVTVLLPGDPAQDRVGVAVYRIAAOSKQAGVLDHLDYRKN
GYRCSLEFHEPTDGAEPQIMVYATQANDSYAGDVQVPCIAQVIFSSAGPSGN
REYLFNLAAMDQLPGFAGVDEHLELVACVKRYIVDEPQLINHALHHEISGLIREQ
LAQAHQLEKLLERCQKPGREWLIYALQAKNKS"
7665..29389
/gene="tomosyn"
/locus_tag="CG17762"
/note="synonyms: CG11056, CG2205; last curated on Mon Nov
04 07:59:47 PST 2002"
/map="11B6-11B7"
/db_xref="FLYBASE:FBgn0030412"
join(7665..7858,9963..10002,11777..12328,12584..12681,
12756..12944,13096..13230,13306..13369,13483..13690,
13756..13827,14064..14218,14389..14570,14638..15120,
15199..15341,16707..16888,27185..27271,27340..27462,
27524..27643,27720..28169,28286..28394,28460..28546,
28614..28709,28784..28908,28992..29389)
/gene="tomosyn"
/locus_tag="CG17762"
/product="CG17762-RB"
/db_xref="FLYBASE:FBgn0030412"
join(9565..10002,11777..12328,12584..12681,12756..12944,

mRNA

mRNA

gene

CDS

mRNA

gene

13096. .13230,13306. .13369,13483. .13690,13752. .13827,
 14064. .14218,14389. .14570,14638. .15120,15199. .15341,
 16707. .16888,27155. .27271,27340. .27462,27524. .27643,
 27720. .28169,28286. .28394,28460. .28546,28614. .28709,
 28784. .28908,28992. .29389)
 /gene="tomosyn"
 /locus_tag="CG17762"
 /product="CG17762-RA"

Alignment Scores:
 Pred. No.: 6,73e-06 Length: 302665
 Score: 743.00 Matches: 237
 Percent Similarity: 41.08% Conservative: 106
 Best Local Similarity: 28.38% Mismatches: 254
 Query Match: 5.62% Indels: 238
 DB: 3 Gaps: 23

US-09-522-753-5 (1-2517) x AE003490 (1-302665)

Qy	144	LysLeuGluProValSerProProSerProProHisThrAspProGluLeuVal	163	Db	88615	GCCTTGGAGGATAAAAGATGCGATCCTATGCGGTTATATACCGCTCTGATGCACGATGCC	88556
Db	89362	CAGGTGGAGGCCATTTCCGCGACATTCGCC-----AGCGATAGTTCCGATTGAGGAGCGC	89309	Qy	403	AspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysVal	422
Qy	164	ProProArgLeuSerLysGluLeuLeuIleGlnAsnMetAspArgValAspArgGluIle	183	Db	89555	CGCCAGAGCGCATGCGCTATCATATGAAATATGGACTTATCGAGGATATGTTGGCAGTT	88496
Db	89308	GGAAGGACCAAGCGCCAAAGAGGATCTCTCATGCAATCCAAAGAGTTGACATGAGATC	89249	Qy	423	TyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGlnLysGluThrPheArgGlu	442
Qy	184	ThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnGluGlu	203	Db	88495	CATCAACAGCGCTAAGGCCCTCAACATGTGGACCGCGCGAGAGGAGACCTTTAAGGAG	88436
Db	89248	AAATCCGCTCAGACGACAAATGGAACGTTGCGCAAAAGAGAAATCCCTCATGGAGGAG	89189	Qy	443	LysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThr	462
Qy	204	Ala-----AlaLysProProGluProGluLysProValSer-----	215	Db	88435	ARGTATTGCGCATCGAAAAAATTTGGAGCCATTTCGCGCTAGTTTGGATCGCAAAATCG	88376
Db	89188	GCCGCCCTGCGCAAGAGCAGAGCGCGCCAAAGAGTTGAACGATAACAATGATCAG	89129	Qy	463	ValAlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeu	482
Qy	216	ProProProIleGluSerLysHisArgSer-----LeuValGlnIleIleTyrAspGlu	233	Db	88375	CCACAGATTGCTTCTACTACTACCTCAGCAGAGNAGACCAGAGATTAACAGCAGCTG	88316
Db	89128	GAACCGATGTGGNACTATCTGTGCGCAGCGCAATGCTAGCGGAGAGATCTATGCGGCC	89069	Qy	483	ValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGln	502
Qy	234	AsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGlu	253	Db	88315	CTTCGAAATCCCGCCCAACGCGCGCAGTTCGCGTAATCCAGCAAGGCTCAGCGCGCA	88256
Db	89068	AATCGGAAGACGCCCAAGCGCAACATTCCTCGTGCAGAAATGCGCGCGCTGATGCTGCAAA	88085	Qy	503	Gln-----GlnGlnGln	506
Qy	254	-----LeuProLeuTyrAsnGlnProSerAspThr	263	Db	88255	CAGCCGCGAGTCATTATCGATTTCGATGACCATCGCGCTTATGACGCGATTCGACGCGAA	88196
Db	89008	TCCTCCGGTTCCGTGCGCGCAGACCATGTTGCCATTGTATAACGACCATCTCGATGC	88949	Qy	507	GlnGlnGlnGlnProMetProArgSerSerGln--GluGluLysAspGluLysGluLys	525
Qy	264	ArgGlnTyrHisGluAsnIleLysIleAsnGlnAla---MetArgLysLysLeuIleLeu	282	Db	88195	CAGCAACAAGATCCCGCGCTGTTTCATCCGCGTGTGCTGAGCGTGAAGTGGCGGACGT	88136
Db	88948	GAGGCACCTTCCCATGCTAATACGCCAGCATCAGATCAAAATTCGTGCGCGCTACTTTTG	88889	Qy	526	GluLysGluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAspLeu	545
Qy	283	TyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyr	302	Db	88135	GCCGCGAGCGTGAGCGTGCGCAGAGAAAGCGCAGCTGATGCTGCAAA-----	88085
Db	88888	CACATAGAAATTTAAGCGCGCAACCGTGGCGGCACAAATCAGGACCTGTGGGAGAGATC	88829	Qy	546	LeuLysGluLysThrAspAspThrSerGlyLysAsnAspGluLysGluAlaValAla	565
Qy	303	AspGlnLeuMetGluAlaLeuGluLysValArgIleGluAsnAsnProArgArg	322	Db	88084	-----GCGGTGAAAGTGCCCGCGGAAAGGACGCGCGCC	88049
Db	88828	ACCAAGGATCAGCGGATTGCGCAGCGAGATGCGGCTATGAGGCGCAGTGCACAGAGG	88769	Qy	566	SerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThrArgSer	585
Qy	323	ArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLys	342	Db	88048	ACTAAG-----	88043
Db	88768	AAAGCAGCGCAGGCAAAATTCGAATCTTCGAAAGGCTCTACGAGGAGTAAAGGAA	88709	Qy	586	MetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAla	605
Qy	343	GlnArgGluLeuGlnGluMetGlnSerArgValGlyGlnArgGlySerGlyLeuSer	362	Db	88042	---GCTGCGAAGCAACGGCT-----	88025
Db	88708	CAGCGGAGGATAAGGAACGTTTC---AATCGGTGCTGCTCGCGG-----	88667	Qy	606	SerMetGluLeuAsnGluSerSerArgTyrThrGluGluGluMetGluThrAlaLysLys	625
Qy	363	MetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGln	382	Db	88025	-----	88025
Db	88666	-----ATTAAATCCGAGCGGATCTCGAGAGATCATGATGCGCTTCGAGGAGCAG	88616	Qy	626	GlyLeuGluHisGlyArgAsnTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIle	665
Qy	383	GluAsnLeuLysGlnMetArgGlnLeuAlaValIleProMetLeuTyrAspAla	402	Db	88025	-----	88025

[illegible][illegible]


```
Qy 854 AspThrGlyysAlaGluGluProValIysSerGluCysThrGluGluAlaGluGluGly 873
Db 150872 CAAGCCAAAGCAGCAGCAACAGCCCA-----AGCCAGCAAGCCAGCCAGCGGCGCCAGCC 150925
Qy 874 ProAlaIysGlyysAspAlaGluAlaGluAlaGluAlaGluAlaGluAlaGluAla 893
Db 150926 AGCCAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 150985
Qy 894 GluIysGlyGluGlyGlySerGlyAlaThrAlaIysSerSerGlyAlaProGln 913
Db 150986 AGCCAGCCCA-----AGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAA 151030
Qy 914 AspSerAspSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAsp 933
Db 151031 GACAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151087
Qy 934 LysAsnArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAla 953
Db 151088 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 151141
Qy 954 AsnAlaSerProGlnIysProLeu---AspLeuIysGlnLeuIysGlnArgAlaAlaAla 972
Db 151142 CGAAGCCAGCCAAATCCACCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151201
Qy 973 IleProProle---GlnValThrLysValHisGluProProArgGluAspAlaAlaPro 991
Db 151202 ACCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151261
Qy 992 ThrIysProAlaProProAlaProProProGlnAsnLeuGlnProGluSerAspAla 1011
Db 151262 GCCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151321
Qy 1012 ProGln-----GlnProGlySerSerProArgGlyIysSerArgSerProAlaPro--- 1028
Db 151322 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151381
Qy 1029 -----ProAlaAspIysGluAlaPheAlaAlaGluAla 1039
Db 151382 GCCAGCCATCAAGCTAGGCAAGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151441
Qy 1040 GlnIysLeuProGlyAspProCysTrpThrSerGlyLeuProPheProValProPro 1059
Db 151442 GCCAAG---CCAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151495
Qy 1060 ArgGluValIleIysAlaSerProHisAlaAspProSerAlaPheSerTyAlaPro 1079
Db 151496 -----AGCCACCAGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151540
Qy 1080 ProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgPro 1099
Db 151541 CCGTCCAAAGCCA-----GCCAAGCCA-----CCAGCCAG 151570
Qy 1100 ProThr-----IleSerAsnProProProLeuIleSerSerAlaIysHisProSerVal 1117
Db 151571 CCAACGAGCTAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151630
Qy 1118 LeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTy 1137
Db 151631 ACAGCCCAAG-----CCAGCC 151645
Qy 1138 SerGluHisAlaIysAlaProValGlyProValThrMetGlyLeuProLeuProMetAsp 1157
Db 151646 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151704
Qy 1158 ProIysIysLeuAlaProPheSerGlyValIysGln----- 1169
Db 151705 ACCCGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151764
Qy 1170 -----GluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyVal 1186
Db 151765 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151824
Qy 1187 ProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValPro----- 1204
```

```
Db 151825 CCAGCCAGCCAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151884
Qy 1205 ---GlyGlySerIleThrIysGlyIleProSerThrArgValProSerAspSerAlaIle 1223
Db 151885 CGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151944
Qy 1224 ThrTyArgGlySerIleThrHisGlyThrProAlaAspValLeuTyIysGlyThrIle 1243
Db 151945 CAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151980
Qy 1244 ThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeu 1263
Db 151981 -----AGCCGCGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152004
Qy 1264 ProIysGlyHisValIleTyThrGluGlyIysGlyHisValLeuSerTyThrGluGlyGly 1283
Db 152005 CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 152061
Qy 1284 MetSerValThrGlnCysSerIysGlu-AspGlyArgSerSerSerGlyPro----- 1300
Db 152062 CTACCAGCCAGCCAGCTGGCCAGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCT 152121
Qy 1301 -----ProHisGluThrAlaAlaProIysArgThrTyAspMetMetGluGlyAr 1317
Db 152122 AGCAAGACACACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 152181
Qy 1317 gValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPr 1337
Db 152182 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152226
Qy 1337 oGluArgHisSerProHisIleIysGluGlnHisIleIleArgGlySerIleThrGl 1357
Db 152227 CAAGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 152278
Qy 1357 nGlyIleProArgSerTyThrValGluAlaGlnGluAspTyThrArgArgGluAlaIysLe 1377
Db 152279 -----CAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152322
Qy 1377 uLeuIysArgGluGlyThrPro-----ProProProProSerA 1391
Db 152323 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152382
Qy 1391 rgAspLeuThrGluAlaTyIys---ThrGlnAlaLeuGlyProLeuIysLeuIysProA 1410
Db 152383 ACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152432
Qy 1410 laHisGluGlyLeuValAlaThrValIysGluAlaGlyArgSerIleHisGluIleProA 1430
Db 152433 -----CCAGTCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 152466
Qy 1430 rgGluGluLeuArgHisThrProGluLeuProLeuAla-----ProArgProLeuL 1447
Db 152467 GCCAGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152526
Qy 1447 ysGluGlySerIleThrGln-----GlyThrProLeuIysTyAspT 1461
Db 152527 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152586
Qy 1461 hrGlyAlaSer-----ThrThrGlySerIysLysHisA 1472
Db 152587 GCCAGCAAGCCAGCTGTCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 152646
Qy 1472 spValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAsp- 1491
Db 152647 AGCCAGCCAGCTGCCAAACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152705
Qy 1492 --ValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyThrGluGluSerLeuIysSerA 1511
Db 152706 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152765
Qy 1511 rgProGlyThrAlaSerSerSerGlyIysSerIleAlaArgGlyAlaProValIleValP 1531
```



```
Db 261385 CCGCGCGCCCTGTCCGGCTGCGCGACCCCTCACCCGGCGCGTGTCTGGAGCGCCCT 261444
Qy 627 -----
Db 261445 GCGCAAGCGCGTAGCGAGCGCCATCTCCGGCCACCGCTACCTGACAGACCGCGCGCG 261504
Qy 628 -----LeuGluHis-----
Db 261505 CCACCTCCCGCAGCAGACGACCTCGCCCTGCTCCAGGGCGTCTTGGCCCTTGCACACAC 261564
Qy 631 -----GlyArgAsnTrpSerAlaIleAlaArgMetValGlySerIysThrValSe 647
Db 261565 CCAGTGGCGGGCGCTGTACCTGTGCGCGCGCAGCAGCCCGCGCTCGCACCTCTAC 261624
Qy 647 rGln---CysLysAsnPheTyrPheAsnTyrIysArgGlnAsnLeuAspGluIleSe 666
Db 261625 CGACCTGTGCGCGACCTGATC-----CGCGCAGCAGGACGGCTCCCA 261669
Qy 666 uGlnGlnHisIysLeuIysMetGluIysGluArgAsnAlaArgArgIysIysIysAl 686
Db 261670 GCGCGGCTCGCGCTGACCCCGCTACGCCACCT-CATCGAGCTGCGCGCCACCCCGACA 261728
Qy 686 aProAlaAlaAsnGluGluAlaAlaPheProProValValGluAspGluMetG1 706
Db 261729 -CCCTGAGAGCTGGCTCTCGAAGGACCGCTCCCGCGCGCGCCGAGACTCGACCCCGAA 261787
Qy 706 uAlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAl 726
Db 261788 CTGGCTGGCGCATCTCTCG-----GCCG 261811
Qy 726 aSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSe 746
Db 261812 CTGCGCTCTCTCGCGCCA-----TCGACGACGACGTCTATC 261847
Qy 746 rAspThrGluSerIleProSerProHisThrGluAlaAlaIysAspThrGlyGlnAsnG1 766
Db 261848 GAGCGCAACTCTGTCAGGACCCAGCGCCAGCGGCGAGAGCGCGCGCTGCCG 261907
Qy 766 yProLysProProAlaThrLeuGlyAlaAspGlyProProPro-----GlyProProTh 784
Db 261908 GCGCGCTGCGCGACCCGAGTCCAGAGCAGCGGCTTGGGAAGAGATGTTTCAACCCGAC 261967
Qy 784 r-----ProProArgArgThrSerArgAlaProIleGluPr 796
Db 261968 CACCTCTCCAATCTACTGTTTACCGCCACCGCCCGCGCTTTCGGCAACCCGAAAGGCC 262027
Qy 796 oThr-----ProAlaSerGluAlaThrGly-----AlaProThrProPro-- 809
Db 262028 GACCTGCTACGGCGCTACGTGACGGGTACTGGACCGCGGTGCGCGTCCGCGCGCGC 262087
Qy 810 -----ProAlaProProSerProSerAl 817
Db 262088 GCGCGCCCGCATCGCGCGCGCGCGCTTGGGCGCTTCCCGCCCGCAGCGCTCAGC 262147
Qy 817 a-----ProProPro-----ValValPro----- 823
Db 262148 CCGGACACCTTCCCGCGCGGGAACAGTGCCTGCGGAGCGCGCCACCCATCCCGGCCCTG 262207
Qy 824 -----LysGluGluIysGluGluThrAlaAlaAlaProPro-----V 837
Db 262208 CGCGCAAACTCGTCGAGCACTCGACGACCTGCGCGGCGCTTCCGGTCCGGGAGGCA 262267
Qy 837 alGluGluGlyGluGluGlnIysProProAlaAlaGluGlu----- 850
Db 262268 TAGCCCCCGGACCGGGGCACACCGCGCGCTCTCCGCGACCGCGCGGTGATCATCCCC 262327
Qy 851 --LeuAlaValAspThrGlyIysAlaGluGluProValLysSerGluCys----- 866
Db 262328 TCCTGGCTCCACCCCGCGCGGGGATGATGATGACCGCGGGGACCCACCGCTAGCACGCC 262387
Qy 867 -----ThrGluGluAlaGluGluGly-ProAlaIysGlyIysAspAlaGluAla 882
```

QY 1191 luAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysG 1211
Db 263369 ACTGCTGTTAGCGCAGCCAGCCAGCCAGCTCCAGCGATCGAACGCGCCACACCGT 263428
QY 1211 lyIleProSerThrArgVal-----ProSerAspSerAlaI 1223
Db 263429 CA---CCCTGAGCTGCACAAACTCGCTGCGCAACCGCGCGCGCTCTCGCGCT 263485
QY 1223 leThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrI 1243
Db 263486 ACGGACCCCGACGACTCGAGCTCCAGCTCCAGCCAGCCGCGACT-----ACCTCAACGC 263539
QY 1243 leThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerL 1263
Db 263540 CGACGACGACACCAAGCGGACTCCCGGACC----- 263571
QY 1263 euProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyG 1283
Db 263571 ----- 263571
QY 1283 lyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerGlyProProHisG 1303
Db 263572 --TCCTCGCGGCTCTCGTGGCACCCAGCCGCGCGCATCTCTAAATCGCGCTCAC 263629
QY 1303 luThr-----AlaAlaProLysArgThrTyrAspMetMetGluG 1316
Db 263630 CCTCAAAACCTCGGCGCGAGCGACTCGGCACCTCGTCGACC----- 263673
QY 1316 lyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleP 1336
Db 263674 -----AGTCTGCGC 263683
QY 1336 roProGluArgHisSer-ProHisHis-----LeuLysGlu 1347
Db 263684 CCAGCCGAGAACTCGCCGCTTCATCCAGGCACCCCGGCTTCGAGCTCTACGACCC 263743
QY 1348 GlnHisHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGln 1367
Db 263744 ACCACCATCAGCAC-----CGTCTGTTCGG----- 263770
QY 1368 GluAspTyrLeuArgArgGluAlaLysLeuLysArgGluGlyThrProProProPro 1387
Db 263771 -----GCCGCGCAAGCCTCCGACGACACCGTCCGCGGCTACGCGCACCTCTCT 263821
QY 1388 -----ProProSerArgAspLeuThrGluAla 1396
Db 263822 CACCGAAGCGCGCGCTCTCGGCGCGCGCTCTCGACGACAGCTCTGGCTCAAGC 263881
QY 1397 TyrLysThrGln-----LeuLysGlu 1400
Db 263882 CACCTTCTCAACCCCGACACCGCGCGGCGGACCTCGCGCGCTCTCTGAAACTGGTGA 263941
QY 1401 -----AlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAla 1416
Db 263942 AGGACACACCCCGATGAGCGCGACGACCCCGCCACCGCCGCGCCAGCAACC-CGA--AGC 263997
QY 1417 ThrValLysGluAlaGly-----ArgSerIleHisGluIleProArgGluGlu--- 1432
Db 263998 ACCCGGCGCTGGCGGGCATCGGCATCGCGCCCTTCAACTCTCTCTCGCGCGCTCTCGC 264057
QY 1433 -----LeuArgHisThrProGluLeuProLeuAla 1442
Db 264058 CCACCCCTCACCGAACTCGACACCGCTTCTACGACAGCGCCCGAGCTTCGACTGGCA 264117
QY 1443 ProArgPro-LeuLysGluGlySer-----IleThr 1452
Db 264118 CCGCGGCTCTCATCGACGCGCGCCACCTCCAAAGTCCCTTTCCTCGCGCGACCTCGTGAC 264177
QY 1452 xGlnGlyThrPro-----LeuLysTyr-AspThrGlyA 1463
Db 264178 CCTCGCGGACCCCAACCGACCCCTGGTCTCTTCTCAACCACTCAAGACCGCGGACCGCT 264237

QY 1463 laSer-----ThrThrGlySerLysLysHisAspValArgSerL 1476
Db 264238 CTTCCCTCTTACTTCCAGCGGCTTCCACTCCAGCGCGCGGCGGATACAGCGCTACTG 264297
QY 1476 euIleGlySerProGlyArgThr-PheProPro----- 1486
Db 264298 CCGTGGGTGCGCGACAGCCTCCCGGCGCTCCACTTCGGCCACCGGTGAGCGCGTCCG 264357
QY 1487 -----ValHisPro-LeuAspValMetAlaAs 1495
Db 264358 CTGGAACCCCGAACCGGACGCTTTCGAAGTCGACTTCACCCACTCGACCGCGCGGA 264417
QY 1495 pAlaArgAlaLeuGluArgAlaCysTyrGluGlySerLeuLysSerArgProGlyThrAl 1515
Db 264418 AGCGAAGCCCTCGGACGACC---TACAGCAAGAACATCGTCTCGCGCTCGGACCGC 264474
QY 1515 aserSerSerGlyGlySerIle-----AlaArgGlyAlaProValIle-V 1530
Db 264475 CCCTAGTCCCGGAACCCCTCAAGCCCTCGTGAAGCCCGCGCTGCCGCTCATCCA 264534
QY 1530 alProGluLeuGlyLys-----ProArgGlnSerProLeuT 1542
Db 264535 CGCGCGGACTACTCCAGCACCGGACCGCTTCTGACCGCGCAACACATCACCGTCAT 264594
QY 1542 hrTyrGluAspHisGlyAlaProPheAlaGlyHis-LeuProArgGlySerProVal--- 1560
Db 264595 CGGCTCAGGACAGTCCGCGCGCG---AAGTCTCTCGACCTGCTCCGCAACCG 264645
QY 1561 -----ThrMetArgGluProThrProArgLeuGlnGlySerLeuSerSerSer 1577
Db 264646 CCGCGCGGACACAGAGGATCCATGGCTCGC---CCGACGAGGCTTCGACCGCAT 264702
QY 1578 LysAlaSerGlnAspArgLysLeuThrSerThrProArg-----GluIleAlaLysSer 1595
Db 264703 GGAGTACTCAAACTCGGCTCGACACTTACCCCGACTACACCGGCTACTTCCACGC 264762
QY 1596 ProHisSerThrValPro----- 1601
Db 264763 CTTCAACGAAACCGTCCGCGACCGGCTCGTCCGCGCCCAATGGCAACTCCACAAAGCAT 264822
QY 1601 ----- 1601
Db 264823 CGACGCGGACACCATCGCGCCATCCACAGAGCTTACCGGCGACCTTCGACGCGG 264882
QY 1602 -----GluHisProHisProLysSerProTyrGlu 1612
Db 264883 CTGCGCGGACGCGCTCTCAGCGCGCGCTCAGCTCCGACCGCGCGCTCGCTCGCCAC 264942
QY 1613 HisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPhe 1632
Db 264943 CAC-----CAAAAGTCGAACTCCA----- 264960
QY 1633 AspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuPro 1652
Db 264961 ---CCTGGAACCGCCAGCGGACCCGCTCCCGCTCCACCAACGCGCTGTCTCT 265017
QY 1653 ArgHisLeuAlaPro---AsnProThrTyrProHis-Leu-----TyrProProTyrLe 1669
Db 265018 CGCCACGCGTACCGGGAACGCGCTCTCGCGCGCATCTCTCGCGGCTCGACCCCTACCT 265077
QY 1669 uIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleLeuAsnAsp-- 1688
Db 265078 GCGCGGACAGCGCGGACCGCGCGCGCATCGCAACACGAGTTCGCGCTCGACCTCGACCC 265137
QY 1689 -----TyrIleThrSerGlnGlnMetHis---HisAsnThrAlaTh 1701
Db 265138 CTGGTACCGGCGCGCTTCTCGAGAACGCGGAACTCCACACCGCATCGCGCAC 265197
QY 1701 rAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAl 1721
Db 265198 C-----CCCGACCTC-----GGCTCGCGCTCGGCGGCGGACCAT 265236
QY 1721 aLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHis-LeuP 1741

Db	265237	CCTCAACTCCCTCACCGGCAAG-----GACCTTACCCTCCACACCGCAC	265284
Qy	1741	roValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrl	1761
Db	265285	CG-----	265286
Qy	1761	euprothrAlaproGlnPropheSerSeraRghHiserSerSerProLeuSerProGlyG	1781
Db	265287	--CCTTCACACCTTCGGACTCCAGCAGCAACACATCCCAACCG--	265331
Qy	1781	lyProThrhIsleuthrLysProThrhTrThrSerSerSerGluArgGluArgAspAra	1801
Db	265332	--CCCCCAGGTCCAG--CCCTCAGCCGCCCTCATCGAGCAGCGCTGCACAACACAGCC	265386
Qy	1801	spArgGluArg-AspArgAspArgGluArgGlu-----LysSer	1813
Db	265387	AACGGCGCGAGAANAACACCGCGAGCCGCTGCACACACCCCTGCGGACCCCGCAGG	265446
Qy	1814	IleLeuThrsThrThrValGluHlsAlaProIleThrArgProGlyThrGluGln	1833
Db	265447	CCCTCGCGGACCCCGCTTGCCCTTTAGAACACCGGGAACCGCTCCCGCTGAGCTCC	265506
Qy	1834	SerSerGlySerSerGlySerSergly-----	1842
Db	265507	AGTCCACCGAGCGAAAGTCTTCGGATCCAGCACACCCCTCGCGTCAACCATCCCGGA	265566
Qy	1843	GlyGlyGlyGlySerSerSer-----ArgProAlaSerHiserHis	1856
Db	265567	TCCGGTCCGGATCTCCGTCGACTCCGACACACAACTCAGCGCCGACGCAATACGGA	265626
Qy	1857	AlaHisGln-----HisSerProIleSerProArgThrGlnAspAlaleu-----	1871
Db	265627	AGGCACGCGCCCGCTTCGCACGTAAGTTGTCACGCGAACAACAGAACTGCTCGCATCAT	265686
Qy	1871	-----	1871
Db	265687	CCAGCGCGCAGCTTGTAAGTCTGAGTCTTGTATCGCGAACCCCGCGCTGCGGATGT	265746
Qy	1872	---GlnGlnArgProSerValLeuHisAasn-----ThrGlyMet	1883
Db	265747	CGATCTCGTACCCAGACCCGACACATAGTCGTAGTTGTAGTCCGCGCGTGTTCGCGT	265806
Qy	1884	LysGlyllelThrAlaValGluProSerLysProThrValLeuArgSerThrSerThr	1903
Db	265807	TCGTGAGTCTTCGCTGTGCACACCGCTGCCCGCGCTGTGCAGAAATACTTCGCGG	265866
Qy	1904	SerSerProValArgProAlaAlaThrPheProProAlaThr-----	1917
Db	265867	AGTACTCAGTAGTCCCGCAGCTGCGCACCCCTGCAGCAACTTCGCGACCGGTGTGT	265926
Qy	1918	HisCysProLeuGlyGlyThrLeuAspGlyValTyProThrLeuMetGluProValLeu	1937
Db	265927	CATACATACAGACTCGACAGATCCCGGATCGTCACTTCGCGCGCGGATCT-----	265980
Qy	1938	LeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHis	1957
Db	265981	---CCGACGTCGGG---AGAACGCGACGCTTCGCGGATCACCGCAACGCGGTACT	266034
Qy	1958	AlaPheLeuAlaLysProProAlaArgSerGlyLeuGluPro-----	1971
Db	266035	GGTGTCCCACCGCGCGCCCTGCAGCATCTCCTGAACCTTGTGTGATCAGATCGATGA	266094
Qy	1972	-----AlaSerSerProSerlysglySerGluProArgProLeuVal	1985
Db	266095	TCGGGCGCTCTGTGAGCGCGCCCTCGACCGCTGTCAGCGTCT---CCGTCCCGTACCGA	266151
Qy	1986	ProProValSerGlyHisAlaThrIleAlaargThrProAlaLysAsnLeuAlaProHis	2005
Db	266152	CCACTGTTTCA---CATACTCA-----	266181
Qy	2006	HlsAlaSerProAsp-----	2010

```
Qy 2300 snGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIle- 2319
Db 267138 AGCGAATACCGCTTCGG---ACGCTTCGGGACC-----CTTCGGCGCCCGCCCTTC 267188
Qy 2320 -----Thrg 2321
Db 267189 ACCTCGCGGCTCGCGCGCGAGCGAGCGCGCGGTCAGCGGACGCCCGCCCG 267248
Qy 2321 lyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetG 2341
Db 267249 GTACGGCGGACTTCTCAAGAACTTCGACGGTTCAACGGCATGTCTCGTACTCTTCGG 267308
Qy 2341 lyLeuGluAlaIleAlaArglysAlaLeuMetGlyIleTyr----- 2354
Db 267309 GAAAGGCCAATGACCGCGGTAGATTCTTGCGCCACAGCCAGCCAGCCCGCCCGC 267368
Qy 2355 -----AspGlnTrpGlu- 2358
Db 267369 AGTTGCGATCTGATGACCACTAGTACCCACCGCCACCGCCAGTGCACAGTGGGACG 267428
Qy 2359 -----GluSerProp 2362
Db 267429 TATGACCCCGACCGCGAGGAACCCCGCCCGCCACAGGCGCCCTACGGACCCCGGAGCC 267488
Qy 2362 roLeuSerAlaAsnAlaPheAsnProLeuAsn-----A 2373
Db 267489 CCGCATCGCGTCCGGCGGAGGACCCACCTCGAAGTCGACCCCGAGATGCGCCCGCATGG 267548
Qy 2373 laSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisr 2393
Db 267549 CATCACCGTCAGCGCCCGCGGCA---CCGACAGACGGGACG-----CACT 267590
Qy 2393 hrLeuThrSerProGlyGly-----GlyGlyLysAlaLysAlaValSerGlyVargp 2409
Db 267591 CACGGACCTCACCGCGCGACACACCGCGGCTCGACTCGTCAAAACCTACGGGGAGC 267650
Qy 2409 roSerSer-----ArgLysAlaLysSerPro----- 2417
Db 267651 CGTCGACGACTGGAAACCGCGGCTTCTCCATCACCCCTGAACTCACCACACGCGCG 267710
Qy 2418 -----AlaProGlyLeuAlaSerGlyAspArgPro----- 2428
Db 267711 CGGCGAAGCATCGCGGCTTACCAACGCGCGCTCCATCACCGCGAACTCACCGACTT 267770
Qy 2429 -----SerValSerSerValH 2434
Db 267771 CACGGCACTGGGGAACCTACGACCGCGGCTGCGGCGACTCGACTCACCGCGTGCACGG 267830
Qy 2434 isSerGluGlyAspCysAsnArgThrProLeuThrAsnArgValTrpGluAsp---- 2452
Db 267831 CCCTGGTGGCCCTCGCGCCGACTCACCGCCACCGCCAGCCCGCGGCAAGCAGT 267890
Qy 2453 -----ArgProSerSerAla---GlySerThrProPheProTyrAsnProLeuIleMetA 2470
Db 267891 GCGCGAAGCGCTCGAGCGCGCCGCAATACGCGCGGCTCGGACACCCCTCGCGCGC 267950
Qy 2470 rgLeuGlnAlaGlyValMetAlaSerPro-ProProGlyLeuProAlaGlySerGly 2489
Db 267951 CCTGTGCAACTGCGGACATCGCGCGGCAACCTCCGCGCTTACCC-----TCGCGC 268004
Qy 2490 ProLeuAlaGlyProHisHisAlaTrpAspGluGluProLysPro 2504
Db 268005 CCGCGCGCATCGCTGCCTCATGCGCTTACGCGCGCGCGCGCC 268048

RESULT 55
LOCUS AR166425/c
DEFINITION Sequence 3 from patent US 6280999.
ACCESSION AR166425
VERSION AR166425.1 GI:16241741
KEYWORDS Unknown.
SOURCE
```

ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 33529)
AUTHORS Gustafsson,C., Betlach,M.C., Ashley,G., Julien,B. and Ziermann,R.
TITLE Sorangium polyketide synthases and encoding DNA therefor
JOURNAL Patent: US 6280999-A 3 28-AUG-2001;
FEATURES
Location/Qualifiers
1..33529
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:
Pred. No.: 2,47e-06 Length: 33529
Score: 726.50 Matches: 674
Percent Similarity: 32.36% Conservative: 314
Best Local Similarity: 22.08% Mismatches: 1199
Query Match: 5.50% Indels: 873
DB: 6 Gaps: 137

US-09-522-753-5 (1-2517) x AR166425 (1-33529)

Qy 17 ProArgTyrProPro---HisSerLeuSerTyrProValGlnIleAlaAArgThrHisThr 35
Db 22686 CCGAGGCTCCACGACACCATCGCTCCCGCGCTCGCACCTTGCA-----TCAGCC 22633
Qy 36 AspValGlyLeuLeuGluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerPro 55
Db 22632 GACCTCGGG-----CGCACACTAGTCTTGGTCGCTCCGCA 22597
Qy 56 GlySerIleIleGlnProGlnArgArgProSerLeuLeuSerGluPheGlnProGly 75
Db 22596 GGCTCAGCACACCGCCACATCGCAGCGCTCAGCTCTCTATCGAGT---GCCCCAGCA 22540
Qy 76 AsnGluArgSerGln-----GluLeuHis----- 83
Db 22539 GCGAGCGGCTCAGCCCCCAGCTCCACTGAGGTACAGCGCCACTTCCAGCGCA 22480
Qy 84 -----LeuArgPro---GluSerHisSerTyrLeuProGluLeuGly 96
Db 22479 CAGCCCGGGTGGTGTTACTTCGCTCGCTCAGCAACGCTCTCTCTCGTGGTCCGCGG 22420
Qy 97 LysSer-----GluMetGluPheIleGlySerLysArgProArgLeuLeuLeu 113
Db 22419 CGCGAACCACCTCTCTCAACCCAGCTGAGATACGCTCCAGCGCTCGCACACCTC 22360
Qy 114 ProAsp-----ProLeuLeuArgProSerPro 122
Db 22359 GTCGAACGCGCGACCGGAACACGGGTACACTTCGTAAAGCTCTTCCCATCCCGAGCCG 22300
Qy 123 LeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThr 142
Db 22299 CTGGTGCCTCG-----CCCGGTGAACAGCAC 22273
Qy 143 GlyLysLeuGluProValSerProProSerProHisThrAspProGluLeuGluLeu 162
Db 22272 CGCAAGCTT---CCCGCTCGCTTCGCGCTCCGCTCAC-----CACCGCGCATC 22225
Qy 163 ValProProArgLeuSerLysGluLeuIleGlnAsnMetAspArgValAspArgGlu 182
Db 22224 CGCGCGCGCGACGACGCGCGGAAGACCTCCACAGCTCCCGCAGCGCTCGCGCGAAG 22165
Qy 183 IleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnGlu 202
Db 22164 CACGAGCGCGGACTCGGAAGTGGCTCGCGTGCAGCGCGCGCT-----GCG 22117
Qy 203 GluAlaAlaLysProProGluProGlu-----LysProValSerProProIle 219
Db 22116 CACCACTCGGACACCCACCTCCCGCTGCTCTTCCAGCACCTTCGCCACCGCCCGC 22057
Qy 220 GluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsn-----ArgLysLys 237
Db 22056 CTGCGCATTCACCGCGGCTCTGTCGACCCGACGACGCGGCAATAGCGAGCGC 21997

Db	2006	ACGTGTTCCGCGCGGTAGTGTCTCTGGCGCGCGCGCAAAAGTACCGCGCGCGCAGC	19947
Qy	788	ArgThrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThr	807
Db	19946	AGAACAGCAACGCGCGAGATCCAGTCCCGCGTCACTGTGTGAGGTGAGCGCCC	19887
Qy	808	ProProProAlaProProSerPro-----SerAlaProProProValValProLys	824
Db	19886	CGTCCACCTTCGGCGCGCAGCACCGCGAGAGCGCTCGGCGCTGTGGCGGTGAGCAGC	19827
Qy	825	GluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGluGluGln---	843
Db	19826	CGTGTGCGACGCGCGCGCGAGTGCAGCACCGCGCTTCAGCGAGCGCGCGTGTGATGC	19767
Qy	844	LysProProAlaAlaGluGluLeuAlaAlaValAspThrGlyLysAlaGluGluProValLys	863
Db	19766	CGCGCAGCACGCGCGACCTCTCCCGCTTCACAGCTCGCAGCAGCACCGTCCACCG	19707
Qy	864	SerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAla	883
Db	19706	TCTCGCGCGCGAGTCTCTCAGTGACTGCACGAGTCTCGCGCGCGCGCTCCAGCC	19647
Qy	884	GluAlaThrAlaGluGlyAlaLeuLysAlaGluLysGlyGlySerGlyArgAla	903
Db	19646	CGC-----GACGCGACGCTCAGCAACAAGTGCCTCACCCCGT---	19611
Qy	904	ThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAla	923
Db	19610	-----GCGCGCGCACGAGTGCAGCGATCGCTGACCGAGTCCCTCCCGCGCG	19557
Qy	924	AspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeu-----Leu	938
Db	19556	TCACGACGACGCGCGCAGGCTCCAGTCCGCGGCTCTCGGTGAGTCTTCCGCTG	19497
Qy	939	SerPro---ArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerPro	957
Db	19496	CCGCTCGCGCGCACGAGCGCGCGCGCACGCGCGCGCGCGCGCGCGCGCGAGTCCG	19437
Qy	958	GlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProIleGln	977
Db	19436	GCTCCGCGC-----CCGTGCGCAGCGCGCGCGCGCGCGCGCGCGCAGC	19398
Qy	978	ValThrLysValHisGluProProArgGluAspAla-----	989
Db	19397	CGTCCAGCGGCTCGTCCCGACATCATGATCAAGCGCGCGCGCTCGGGTGTCTCGCTGC	19338
Qy	990	AlaProThrLysProAlaProAla-----ProProProProGlnAsnLeuGlnPro	1007
Db	19337	GCGCCGCGGAACAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCCG	19278
Qy	1008	GluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAla	1027
Db	19277	GCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT	19218
Qy	1028	ProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProPro	1047
Db	19217	CCGACAGCCAGCTTCAGCAGCAGCAGCG-----CCTGCGCTCGTCCGCT	19173
Qy	1048	CysTrpThrSerGlyLeuProPheProValProProArgGluValLysAlaSerPro	1067
Db	19172	-----CGTGGGCAACAGCGCGCACCTCCGACTGACTCTGTCTCTAAAGTGTGTGT	19122
Qy	1068	HisAlaProAspProSerAlaPheSerTyrAlaProGly-----	1081
Db	19121	CGACCGCGCGCGCGCGCGCGCTCGCGCTCTCGAGCGCGCGCAACCAATGCTTGA	19062
Qy	1082	-----HisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArg	1098
Db	19061	GATCGGCAATCGCTCGCGCTCAGGGCT-----CGGCGAGCTCGCTCGCTC	19014
Qy	1099	ProProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeu	1118

QY	1402	euGlyProLeuLysLeu-Lys-----ProAlaHisGluGlyLeu 1414	QY	1693	lnGlnMethHisAsnThrAlaThrAlaMetAlaGlnAlaArgAspMetLeuArgGlyL 1713
Db	18039	TGAGCCCCGAGGCTCAGCTCGCTACGGCGCTTCGAGTATCCAGCCAGTAAGCTGCC 17980	Db	17037	-----CCGCTCCCGAAGCGCGCTCTGCTCCAGAGCGCGCGCC 16995
QY	1415	ValAlaThrVallysglualaGlyArgSerIleHisGluIleProArgGluLeuArg 1434	QY	1713	euSerProArgGluSerSerLeuAlaLeuAsnTyrAlaalaGlyProArgGlyIleAla 1733
Db	17979	GCTGAACCGGTACGTCGCGAGCT-----CCAGCACAGAC---CGC 17941	Db	16994	CATCCCGCGCACCTGGCTCCCTT-----GGCCGGGGAACACGAACACCACTT 16947
QY	1435	HisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGly 1454	QY	1733	spLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAla 1753
Db	17940	CGTGGCCCGACAGACCCGCTGCTCAGTCCACCT-----CGT 17905	Db	16946	GCCTCGCGCAGCGCGCTACCCCGC-----ACACCGCCGGTGGCGCGACCTCGCA 16893
QY	1455	ThrProLeuLysTyrAspThrGlyAlaSerThrGlySerLysHisAspValArg 1474	QY	1753	hrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHis 1773
Db	17904	GCCCCGTGACGTGCGCGTCCAGCGTCTGTGTACAGCTGCGACACCCAC---CTTGT 17848	Db	16892	CAGG-----CCCGCAGCGCTCTCCGCTCGACACGCTCGC 16854
QY	1475	SerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAla 1494	QY	1773	erSerSerProLeu-----SerProGlyGlyProThrHisLeuThrL 1787
Db	17847	CGCGTGCAGGCTCCCCACACCA-----CCCCCTCGCCTCCCGCAGCGCTGCTCA 17794	Db	16853	CGCAAGCACCATGCGGTGAGCGAAGTGTCTCGGTGAGCGCCCTGTCTCTCACCAC 16794
QY	1495	AspAlaArgAlaLeuGluAla---CysTyrGluGluSerLeuLysSerArgProGly 1513	QY	1787	ysProThrThrThrSerSerGluArgGluArgAspArgAspArgGluArgAspArgA 1807
Db	17793	CGCGCATCGCAGACCGGTGCGCGTCCAGTCCAGACACACCGTCCCTCTCA 17734	Db	16793	GTCCGACCATGCACCTCCCGTCTCTCCGAGCCACTTCCGCCCGCTCCGCTCGC 16734
QY	1514	ThrAlaSerSerSerGlySerIleAla-Arg-----GlyAlaProValL 1529	QY	1807	spArgGluArgGluLysSerIleLeuThrSerThrThrThrValGluHisAlaProL 1827
Db	17733	GCAGCTTCGACGCGCTCGTCCAGGCGCACCGTCCGCCGAAGGTTCGCGCACAGTACT 17674	Db	16733	CCCCACCGAGGCTCGTCTCCCGCCGACAGCAACAGCGGCATCCGACACCTCCGTCG 16674
QY	1529	eValProLeuGlyLysProArgGln-SerProLeuThrTyrGluAspHisGly--- 1547	QY	1827	tpArgProGlyThrGluGlnSerSerGlySerSerGlySerGlyGlyGlyGlyGly 1846
Db	17673	CGCGTCCAGCGCTCCCGCGCACACCTCTCTGTCACTGTCGAGTAAACGGCAGCT 17614	Db	16673	C-AACGGCTCGCGCATTCCTCCACCGCTCGCGCGCGCTCCACCGCGCTCTTCA 16615
QY	1548	-----AlaProPheAlaGlyHis----- 1553	QY	1847	-----SerSerSerArgProAlaSerHisSerHisAlaHisGlnHisSerPro 1862
Db	17613	CGCTCGCTTCGCGTTCGCGACGACGACGCTTCGCTCTAGTCTCGGCGAGCGCATCCA 17554	Db	16614	GGATGCATGCGCGT-CGTCCGCTGATCCCGAACGACGACGCGCGCGCGCGCGCC 16556
QY	1554	-----LeuProArgGlySerProValThrMetArgGluProThrProArgLeuG 1570	QY	1863	-----IleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeu-His 1879
Db	17553	TGTGCGCGCTGCGCGCGTAGTTCAGCTTCCTTCGCGCAGACACACCTTCTCGCG 17494	Db	16555	CGCGCGCTTCGCGCGCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16496
QY	1570	lnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrPro 1590	QY	1880	---AsnThrGlyMetLysGlyIleThrAlaValGluProSerLysProThrValLeu 1898
Db	17493	TCAGTCCACCATAGCCGCTCCAGCGCTCCAGCTCCAGTCCAGTCCAGTCCAGTCC 17435	Db	16495	CAATGTGCGGCTCGGCTCGCTCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCG 16436
QY	1590	tgGluIleAlaLysSer----- 1595	QY	1899	ArgSerThrSerSerSerProValArgProAlaAlaThrPheProProAlaThrHis 1918
Db	17434	GAGGTTCACGCGCTATCGCAAGCGCTCCCGTACGCGCGCATCGCTCTCGCAC 17375	Db	16435	GACAGCACCATCTTGAGCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16388
QY	1596	---ProHisSerThrValProGluHisHis-----ProHisProI 1608	QY	1919	CysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeu 1938
Db	17374	TCCGACACCGCGCTCGACAGCATCATCGCCCGCCATCCCGCGCGCTCGCGCGCG 17315	Db	16387	TGTCGA---GGTTCGACTTCGAGCGCAAGTACAGCGCGCGCTCGCGCTTCGCGCGCG 16331
QY	1608	leSerProTyrGluHisLeuArgGlyVal-----SerGlyValAspLeuT 1624	QY	1939	ProLysGluAlaProArgValAlaArgProGluArg----- 1950
Db	17314	TGGTTCGCGAGCGCTACCGCGCTCTCTCCGCAAGCGACAGCGCTTCGCGCACAC 17255	Db	16330	GCCCGAACACCTCCAGCGCTTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCG 16271
QY	1624	yrArgSerHisIleProLeuAlaPheAspPro-----T 1635	QY	1951	-----ProArgAlaAspThrGly-----HisAlaPheLeu 1960
Db	17254	GCGCGCGACACCTCGCGCTGGGTGT-GGCCACACCGCGCGCGCGCTCGCGCGCG 17196	Db	16270	CATCGCGCTCCACCGCTGATGTCTCGGCGCAAGCGCGCGCGCGCGCGCGCGCGCT 16211
QY	1635	hrSerileProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProArgHisL 1655	QY	1961	AlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGlu 1980
Db	17195	CGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17136	Db	16210	GGATCACCGCTGCTGGCGAGCGCGTTCGCGCGCTCAACCTCGCTCGCGCTCGCGCT 16151
QY	1655	euAlaProAsnProThrTyrPro-----HisLeuTyrProTyrTyrLeuIleArgGlyT 1673	QY	1981	ProArgProLeuVal-----ProProValSerGlyHis----- 1991
Db	17135	CGCGTCCAGCGCGCTGCTCTCCCGCGCATCTCCGCGCGCGCGCGCGCGCGCGCG 17076	Db	16150	GGTTCACCGCTGACCCACCGGATCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16091
QY	1673	yrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerg 1693	QY	1992	-----AlaThrIleAlaArgThrPro-----AlaLysAsnLeuAla 2003
Db	17075	GCCGCTCCAGCGCGCGCGCGCTCATCGCACGCGCTCA----- 17038	Db	16090	ACAGCGCTTCAGCACCGGATCCCGCACCCCTCGGACCGAGATGACGCGCTGAGTCGCG 16031

Db 16030 CAGAGAACTCTTTGCAACGACCGTCCGTCGCAAGGGCTTTGAGCGGCTGAACCTCCACGA 15971
Qy 2004 -----ProHisHisAlaSerProAspProAlaPro 2014
Db 15970 ATCCCGGGGTGGTTCATCATCATCATCCCGTCACCGGTCAAGGCAGATCGCACTCGGCCCT 15911
Qy 2015 ProAla-----SerAlaSerAspProHisArgGluLysThrGlnSerLys 2029
Db 15910 GAGCAGGGCGGTGCACCGCCAGGTGCAGCAGACACACGACGACGACGACGACGCGGTGTCCA 15851
Qy 2030 ProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSer 2049
Db 15850 CCG-----TGATCGCTGGGCGCTTCGAGCCCGGACGACGACGTACG 15815
Qy 2050 ProGluGlyValGluProValSerProValSer---SerProSerLeuThrHisAspLys 2068
Db 15814 CCACAGCGCTGAGACGACGCTCGCCCGGTCCCGGTGCCCGGT----- 15770
Qy 2069 GlyLeuProLysHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArgPro 2088
Db 15769 AGCGTCCAAAGCGCGCAGGTGCGTCCGTAAGACGACCGTAGTCCGAGC-----CCA 15716
Qy 2089 LysGlnProGlyProValLysLeuGlyGluAlaAlaHisLeuProHisLeuArgPro 2108
Db 15715 TCGAGCCAGGTACACCCCGGTGGAGCTCTCGCTCAGCGCGCAGCGGCGCAGCGCC 15656
Qy 2109 LeuProGluSerGlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGly 2128
Db 15655 GCTCCA---GTGCTCCACGCGCTCTCCA-----GCACGCGCGCTGCTGGG 15611
Qy 2129 HisGlnArgValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThr 2148
Db 15610 GATCCATCGCTCGCGCTCTCGGGCG-----ATATCC 15578
Qy 2149 ArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGly 2168
Db 15577 CGAAGAAACCGCATCGA---ACAGGTGCGATGTCCCGCAGGAATCCACCCCTCGCGCACGT 15521
Qy 2169 AlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProPro 2188
Db 15520 AGTCTTGCCCA-----CCGCTCCG 15500
Qy 2189 AspHisGlyAlaProAlaArgGlySerProHisSerGluGlyLysArg----- 2205
Db 15499 GATCGGGGTGTAATCGAAGTGCATCCAGCGCGCAGCGAGGCGCTCGATCGCATCTC 15440
Qy 2206 SerProGluProAnLysThrSer----- 2213
Db 15439 TCCCTCTCTCCAGGACGCGCAGTAGTCTTCCGCGGTAGCGAGCGCCCGCGCGCGC 15380
Qy 2214 -----ValLeuGlyGlyGlyGluAspGlyIleGluProVal 2225
Db 15379 ACCGCATCGATAGTGGCGATGGCGCTCAAGCGAGATCCGAGTGGCTGGCTGCAGCC 15320
Qy 2226 SerProGluGlyMet-----ThrGluProGlyHisSer----- 2237
Db 15319 TCTGCAACTCTTCAATGCAAGGTCACCGCTCAAGAGCGTAGGAGCGCGCGTGC 15260
Qy 2238 -----ArgSerAlaValTyrProLeuLeuTyrArg-----Asp 2248
Db 15259 TATTCATGGAGATCTCTGAGTTCGAAATCATTTGCAAGGCTTGCAGGCGCAGCTGCTCG 15200
Qy 2249 GlyGluGlnThrGluProSerArgMetGlySer----- 2259
Db 15199 TCGCGGAGATCTTCGAAATCGAAATCGAAATCGAAACCTCGAGCGATGTTTCGAGCGC 15140
Qy 2260 ---LysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSer 2278
Db 15139 CCCCGCAGCTCCAAGACCTTGGGGAGGAGGCTGAACGACGACGAGACTGACGAGGAGG 15080
Qy 2279 AsnSerAlaMetValLysSerLysLysGlnGluLeuAsnLysLysLeuAsnThrHisAsn 2298
Db 15079 CTGAGCATGAACCGCGGAATTTCTCGTCGCGCATGAGTCCGCTGTACGACGACGATTC 15020
Qy 2299 ArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAla 2318
Db 15019 CGCGCTCACCGTTTCGAACCTTTGACAGAGGAC-----TTCCGATGTCGTG 14969
Qy 2319 IleThrGly-----ThrGlyLeuMetThrTyrArgSerGlnAlaVal 2332
Db 14968 GCAGATGATGCTCGAAAGCAAAAGTGGCAGCAAGTTGCTTCCGACGACGACGAGCGAGC 14909
Qy 2333 GlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGly 2352
Db 14908 CGATTCGCGACTTCGACGGCCATG----- 14885
Qy 2353 LysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsn 2372
Db 14884 -----ACGGAATTCATTCGAGCTCCCTCAAGGCTGGCGGTAGGATTC 14840
Qy 2373 AlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHis 2392
Db 14839 GCGTCA-----CCGCGCTGAAGCCGACGACCGCGGACCTCGGCCGACCAAGG 14789
Qy 2393 ThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArg 2412
Db 14788 TCATGAGCGAGCTC-----AGCGCTCCGCTCCGCGACGCGCGAGAGCGC 14741
Qy 2413 LysAlaLysSerProAlaProGlyLeuAlaSerGlyAsp-----ArgProProSerVal 2430
Db 14740 TCGCGGAGCGTCCAGCGCTCTCTCTCGCGGACGAGACGCTTGCAGCGCCCGGGCGC 14681
Qy 2431 SerSerValHisSerGluGly-----AspCys 2439
Db 14680 ACCAGCGCGAAGACGCGCGCGCTCGCGCTGGACTCCAGCCCGCGCTGCAGCTGC 14621
Qy 2440 AsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySer 2459
Db 14620 GCGAGATCCAGGTGCATCGGACGACGCTCGCTTCAGAGCGTTCAGAGCGCGGTCCAGC 14561
Qy 2460 ThrProPheProTyrAsn-ProLeu-----IleMe 2469
Db 14560 AGAGCGAGCGCTCGTCGACCGATATCGGACGCGCGCTCGCGCTGTATGCGCGACAGT 14501
Qy 2469 ThrGluGlnAlaGlyValMetAlaSerProProPro-----ProGlyLeuProAla-- 2486
Db 14500 TCGCGCTCGCGCAGGTGCGCTGTCATCCCGCAGCGCGCTTGGCGCCCAACCCCGCGC 14441
Qy 2487 -GlySerGlyProLeuAla 2492
Db 14440 AGGCTCGTCGCGCGAGCC 14422
RESULT 56
AC114730/C 193894 bp DNA linear PRI 02-NOV-2002
LOCUS Homo sapiens BAC clone RP13-925E23 from 2, complete sequence.
DEFINITION AC114730
ACCESSION AC114730
VERSION AC114730.5 GI:23463016
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 193894)
AUTHORS Tomlinson,C., Bielicki,L. and Doebber,A.
TITLE The sequence of Homo sapiens BAC clone RP13-925E23
JOURNAL Unpublished (2001)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 193894)
AUTHORS Tomlinson,C., Bielicki,L. and Doebber,A.
TITLE The sequence of Homo sapiens BAC clone RP13-925E23
JOURNAL Unpublished (2001)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 3 (bases 1 to 193894)
AUTHORS Waterston,R.H.

Mon Apr 19 08:15:02 2004

TITLE
JOURNAL
Submitted (11-MAR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE
AUTHORS
4 (bases 1 to 193894)
Waterston, R.H.
TITLE
JOURNAL
Submitted (07-MAY-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE
AUTHORS
5 (bases 1 to 193894)
Waterston, R.H.
TITLE
JOURNAL
Submitted (02-OCT-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE
AUTHORS
6 (bases 1 to 193894)
Waterston, R.
TITLE
JOURNAL
Submitted (02-NOV-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 2, 2002 this sequence version replaced gi:20486437.

COMMENT

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu

Summary Statistics

Center project name: H_FH0925E23

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPCI-13 Human Female BAC Library was constructed using improved cloning techniques developed by Kazutoyo Osoegawa. The library was generated by Baohui Zhao in our laboratory. Construction was funded by a grant from the National Human Genome Research Institute (NHGRI, NIH) (H1R01RG01165-03). The library was generated according to the new NHGRI/DOE 'Guidance on Human Subjects in Large-Scale DNA Sequencing'.
Female blood was obtained via a double-blind selection protocol. Female blood DNA was isolated from one randomly chosen donor (out of 10 female donors) and partially digested with a combination of EcoRI and EcoRV. The library was generated according to the standard protocol for library segments 3k4. Size selected DNA was cloned into the pBAC3.6 vector between the EcoRI sites for library segments 1k2 or the BamHI sites for library segments 3k4. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). The library has been arrayed into 384-well microtiter dishes and also gridded onto 2x22cm nylon high density filters for screening by probe hybridization.

NEIGHBORING SEQUENCE INFORMATION:
There is an unresolved tandem repeat from base 25865 to 27600.

There is an unresolved tandem repeat from base 39010 to 43450.
There is an unresolved tandem repeat from base 81306 to 83536.
Polymorphisms have been identified between AC133781 and this sequence.

Data from AC133781 was used to finish this clone.

This sequence is the entire insert of the clone.

FEATURES	source	Location/Qualifiers
repeat_region	1..193894	/organism="Homo sapiens"
repeat_region	306..414	/mol_type="genomic DNA"
repeat_region	543..673	/db_xref="taxon:9606"
repeat_region	1274..1526	/chromosome="2"
repeat_region	1707..1727	/map="2"
repeat_region	1987..2074	/clone="RP13-925E23"
repeat_region	2856..3211	/clone_lib="RPCI-13"
repeat_region	3416..3453	/rpt_family="Alu"
repeat_region	3454..3759	/rpt_family="L1"
repeat_region	3760..4103	/rpt_family="L1"
repeat_region	4325..4773	/rpt_family="ERV1"
repeat_region	4778..4982	/rpt_family="ERV1"
repeat_region	4983..5280	/rpt_family="ERV1"
repeat_region	5281..5394	/rpt_family="Alu"
repeat_region	5395..5691	/rpt_family="Alu"
repeat_region	5692..5874	/rpt_family="Alu"
repeat_region	5875..6007	/rpt_family="Alu"
repeat_region	6008..6279	/rpt_family="ERV1"
repeat_region	6280..6398	/rpt_family="Alu"
repeat_region	6399..6689	/rpt_family="ERV1"
repeat_region	6690..6781	/rpt_family="Alu"
repeat_region	6782..7072	/rpt_family="ERV1"
repeat_region	7073..7343	/rpt_family="Alu"
repeat_region	7408..7818	/rpt_family="ERV1"
repeat_region	7890..7986	/rpt_family="Alu"
repeat_region	7987..8027	/rpt_family="Alu"
repeat_region	8073..8220	/rpt_family="ERV1"


```
Db 42359 --CCCTCCCTCCCTCAGCCAGGACCTCCGAGGACACCTCCCTCCCTCCCTCAGCCAGG 42302
Qy 1111 SerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSer 1130
Db 42301 GACCTCCGAGGACACCTCCCTC-----CCCTCAGCCAGGACCTCT 42257
Qy 1131 ValGlnLeu---HisValProTyrSerGluHisAlaLysAla-----ProValGlyPro 1147
Db 42256 GCAGGCTCCCTCCCTCCCTCCCTCAGCCAGGACCTCCGAGGACACCTCCCTCCCTCC 42197
Qy 1148 ValThrMetGlyLeuProLeuProMetAspProLysLeuAlaPro-----PheSer 1165
Db 42196 TCACCCAGGAGCCTCCGAGGACAC--CTCTCTCCCTCCCTCAGCCAGGACCTTCGCA 42138
Qy 1166 GlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGly----- 1179
Db 42137 GGCACCTCCCTCCCTCCCTCAACCCAGGACCTCCGAGGACACCTTCCTCCCTCCCTC 42078
Qy 1180 ---ProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThr 1198
Db 42077 AACCCAGGACCTCCGAGGACACCT-----CCCTCCCTCCCTCAGCCAGGAC 42027
Qy 1199 AlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValPro 1218
Db 42026 TCAGGAGGACCTCCCTCCCTCA--ACCAGGACCTCCGAGGACACCTCCCT 41970
Qy 1219 SerAspSerAlaIle---ThrTyrArgGlySerIleThrHisGlyThrProAlaAspVal 1237
Db 41969 CCCCCCTCAGCCAGGACCTTCGAGGCTCCCTCCCTCCCTCCCTCAGCCAGGACCTC 41910
Qy 1238 LeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArg 1257
Db 41909 CGCAGGACCTCCCTC----- 41871
Qy 1258 GlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysGlyHisVal 1277
Db 41870 CGCAGGCTCCCTCCCTCCCTC-----CTCAGCCAGG 41838
Qy 1278 LeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSer 1297
Db 41837 -----ACCTCCGAGGACACCTCCCTCCCTCCCTCAGCC 41805
Qy 1298 SerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArg 1317
Db 41804 CAGGACCTCCGAGGACCTCCCTCCCTCCCTCAGCC-----CAGGACCT 41757
Qy 1318 ValGlyArgAlaIleSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPro 1337
Db 41756 CTGAGGACCTCCCTCCCTCCCTCAACCCAGGACCTCTGAGGACACCTCCCTCCCTCC 41697
Qy 1338 GluArgHisSer-ProHisHisLeuLysGluGlnHisIleArgGlySerIleThrG1 1357
Db 41696 CCC-----TCAGCCAGGACCTCCGAGG----- 41672
Qy 1357 nGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLe 1377
Db 41671 -----CAACCTCCCTCCCTCCCTCAGCCAGGAC---CTCCGAGGACACCTCCCT 41622
Qy 1377 u-----LeuLysArgGlyThrProProProProProProProProProPro 1391
Db 41621 CCCCCCTCAGCCAGGACCTCCGAGGACACCTCCCTCCCTCCCTTAAGCCC-----AG 41568
Qy 1391 GAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHi 1411
Db 41567 GACCTCCGAGG-----CAACCTCCCTCCCTCCCTCAG-- 41534
Qy 1411 sGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgG1 1431
Db 41533 -----CCAGGGA 41526
Qy 1431 uGluLeuArg-HisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySer1 1451
Db 41525 CCTCTGAGGACACCTCCCTCCCTCCCTCCCTCAGCCAGGACCTCCGAGG----- 41473
```

```
Qy 1451 leThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysH 1471
Db 41472 --CCACCTCCCTCCCTCAGCCAGGACCTCCGAGGACCTCCCTCCCTCCCTCCCTCC 41419
Qy 1471 isAspValArgSerLeuIleGlySerProGlyArgThrPhePro-ProValHisProLeu 1490
Db 41418 -----TCAGCCAGGACCTCCGAGGACCTCCCTCCCTCCCTCCCTCCCTCCAG 41370
Qy 1491 AspValMet-----AlaAspAlaArgAlaLeuGluArgAlaCysTyr 1504
Db 41369 GACCTCCGAGGACACCTCCCTCCCTCCCTCAGCCAGGACCTC-----TCGAGG 41319
Qy 1505 GluGluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArg 1524
Db 41318 CCACCTCCCTCCCTCAGCCAGGACCTCCGCA----- 41283
Qy 1525 GlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGlu 1544
Db 41282 GGCCTCCCT-----CCCTCCCTCCCTC----- 41262
Qy 1545 AspHisGlyAlaProPheAlaGlyHisLeuProArgGlySer-----ProValThr 1561
Db 41261 ---AACCCAGGACCTCCGAGGACACCTCCCTCCCTCAGCCAGGACCTCTGAG 41205
Qy 1562 MetArgGluProThrProArgLeuGlnGlySerLeuSerSerSerLysAlaSerGln 1581
Db 41204 GCCTCCCTCCCTCCCTCAG----- 41184
Qy 1582 AspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHis-Ser----- 1598
Db 41183 -----CCAGGACCTCCGAGGACCTCCCTCCCTCCCTCCCTCCCTCCCTCCAG 41145
Qy 1599 -----ThrValProGluHisProHisProIleSerProTyrGluHisLeuLeuArg 1616
Db 41144 CCAGGACCTCCGAGGACACCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCAG 41102
Qy 1616 gGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSe 1636
Db 41101 -GGGACCTCCGCA-----GGCACCTCCCTCCCTCCCTCCCTCCCTCCCTCCAG 41063
Qy 1636 rIleProArgGlyIleProLeuAspAlaAlaIleTyrTyrLeuProArgHisLeuAl 1656
Db 41062 -----GGACCTCCGAGGACACCTCCCTCCCTCCCTCCCTCCCTCCCTCCAG 41028
Qy 1656 aProAsnProThrTyrProHis-LeuTyrProProTyrLeuIleArgGlyTyrProAsp 1676
Db 41027 CCAGGACCTCCGAGGACACCTCCCTCCCTCCCT----- 40993
Qy 1676 hrAlaAlaLeuGluAsnArgGlnThrIleAsnAspTyrIleThrSerGlnGlnMetH 1696
Db 40993 ----- 40993
Qy 1696 isHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArg-----GlyLeuS 1714
Db 40992 -----CAGCCAGGACCTCCGAGGACCTCCCTCCCTCCCTCCCTCCCTCC 40959
Qy 1714 erProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleAspL 1734
Db 40958 CTCCCTCCCTCAG-----CCAGG-----GACC 40938
Qy 1734 euSerGlnValProHisLeuProValLeuVal-----ProProThrP 1748
Db 40937 TCAGGAGGACCTCCCTCCCTCCCTCAACCCAGGACCTCCGAGGACCTCCCTCCCTCCCTCC 40878
Qy 1748 roGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnPro- 1767
Db 40877 CCCTCAACCCAGGACCTCT-----CGAGGACCTCCCTCCCTCCCTCCCTCAACCCAG 40827
Qy 1768 --PheSerSerArgHisSerSerSerProLeuSerProGly-----GlyP 1782
Db 40826 GGACCTCCGAGGACACCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 40767
```

QY 1782 roThrHisLeuThrLysProThrThrThr-SerSerSerGluArgGluArgAspArgAsp 1801
Db 40766 CTTCCCTCCCTCAGCAAGGAGCTGTGAGGACCTTCCCTCCCTCAACCCAGGAGC 40707
QY 1802 ArgGluArg-AspArgAspArgGluArgGluLysSerIle 1814
Db 40706 CTCGCGAGCCACCTCTCCCTCCCTCAGCCAGGAGCTCCGAGGACCTTCCCTC 40647
QY 1815 --LeuThrSerThrThrValGluHisAlaProIleThrArgProGlyThrGluGln 1833
Db 40646 CCCCCTCAACCCAGGAGCTCGCGAGCCACCTCTCC----- 40611
QY 1834 SerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAlaSer 1853
Db 40610 TCCCCCTCAGCCAGGAGCTTCGCGAGCCACCTCTCCCTCCCTCAGCCAGGAGC 40551
QY 1854 HisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAla----- 1870
Db 40550 ---TCTGAGGACCTCTCCCTCCCTCAGCCAGGAGCTTCGCGAGGACCTCTCC 40494
QY 1871 -----LeuGlnGlnArgProSerValLeuHisAsnThrGly 1882
Db 40493 TCCCTCCCTCAGCCAGGAGCTTCGCGAGGACCTTCCCTCCCTCAGCCAGGAG 40434
QY 1883 MetLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSer 1902
Db 40433 -----ACCTCGCAGGAGCCCTCTCCCTCCCTCC-----TCA 40404
QY 1903 ThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGly 1922
Db 40403 ACCCAGGAGCT-----CCGAGGAGCTCTCCCTCCCTCCCTCCCTCAGCCCA 40356
QY 1923 GlyThrLeuAspGlyValTyProThrLeuMetGluProValLeuLeuProLysGluAla 1942
Db 40355 GGGACCTCCGAGGAGCTCTCCCTCCCTCAGCCAGGAGC-CTC---CGAGGCT 40300
QY 1943 ProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLys 1962
Db 40299 CCCCCTCTC-----CCCTCAGCCAGGAGCTTCGCGAGGAGC----- 40261
QY 1963 ProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArg 1982
Db 40260 CTTCCCT-----TCCCTCCCTCAGCCAGGAGC-----TCCGAGGAGCTTCCCTCC 40213
QY 1983 ProLeuValProProValSer---GlyHisAlaThrIleAlaArgThrProAlaLysAsn 2001
Db 40212 CCCCCTCAGCCAGGAGCTTCGCGAGGAGCCTCTCCCTCCCTCAACCCAGGAGCTCC 40153
QY 2002 LeuAlaProHisHisAlaSerProAspProPro-AlaProProAlaSerAla-SerAsp 2021
Db 40152 ---GCAGGCCACCTCTCCCTCCCTCAGCCAGGAGCTTCGCGAGGAGCTTCCCTCC 40096
QY 2021 roHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuArgSerL 2041
Db 40095 CCCCCTCAACCCAGGAGCTTCGCGAGGAGCCTCTCCCT----- 40058
QY 2041 euGlyTyHisGlySerSerTySerProGluGlyValGluProValSerProValSer 2061
Db 40057 -----CCCTCAGCCAGGAGCT 40039
QY 2061 erProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerH 2081
Db 40038 CC-----GCAGGCC 40030
QY 2081 isLeuGluGlyGluLeuArgPro-----LysGlnProGlyProValLysL 2096
Db 40029 ACCCTCCCTCCCTCAACCCAGGAGCTTCGCGAGGAGCTTCCCTCCCTCCCTCAAC 39970
QY 2096 euGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSer 2116
Db 39969 CAGGAGCTTCGCGAGGAGCTTCC-----TCCCTCCCTCAGCCAGGAGCTTCGCGAG 39916
QY 2116 erSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuA 2136

Db 39915 GCCACCTCTCCCTCCCTCCCTCAGCCAGG-----ACCTCCG 39880
QY 2136 laGlnHisIleSerGluValIleThrGlnAspTyThrArgHisHisProGlnGlnLeuS 2156
Db 39879 CAGGCCACCTTCCCTCCCTCCCTCAACCCAGGAG-----CTCGGAGGAGC 39835
QY 2156 erAlaProLeuProAla-ProLeuTy-SerPheProGlyAlaSerCysProValLeuAsp 2175
Db 39834 CCCCCTCCCTCCCTCAGCC-----AGGAGC 39808
QY 2176 LeuArgArgProProSerAspLeuTyLeuProProProAspHisGlyAlaProAlaArg 2195
Db 39807 CTCGCGAGGAGCTCTCCCTCCCTCAGCCAGGAGCTTCGCGAGGAGCTTCCCTCC 39751
QY 2196 GlySerProHisSerGlyGlyLysArgSerProGluProAsnLysThrSerValLeu 2215
Db 39750 ---TCCCTCCCTCAACCCAGGAGCTTCGCGAGGAGCTTCCCTCC-----TCCCTCCCTC 39703
QY 2216 GlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGly 2235
Db 39702 AACCCAGGAGCTTCGCGAGGAGCTTCCCTCCCTCCCTCAGCCAGGAGCTTTCGAGG 39643
QY 2236 HisSerArgSerAlaValTyProLeuLeuTyArgAspGlyGluGlnThrGluProSer 2255
Db 39642 CACCTTCCCTCC-----CCCTCAGCCAGGAGCTTCGCGAGGAGCTTCCCTCC 39592
QY 2256 ArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPheSerLysLeu 2275
Db 39591 CCCCCTC-----AGCCAGGAGCTTCTGC-AGGCCACCC----- 39560
QY 2276 ThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLysLysLeuAsn 2295
Db 39560 ----- 39560
QY 2296 ThrHisAsnArgAsnGluProGluTyAsnIleSerGlnProGlyThrGluIlePheAsn 2315
Db 39559 -----TCC 39557
QY 2316 MetProAlaIleThrGlyThrGlyLeuMetThrTyArgSerGlnAlaValGlnGluHis 2335
Db 39556 CTTCCCTCTCCCTCAGCCAGGAGCTGTCTG-----CAGGACCTTCTCCCTCC 39509
QY 2336 AlaSerThrAsnMetGlyLeuGluAlaIleArgLysAlaLeuMetGlyLysTyArg 2355
Db 39508 CCCCACCCAGGAGCTTCGCGAGGAGCTTCCCTCC----- 39476
QY 2356 GlnTrpGluGluSerProProLeuSerAlaAsn-----AlaPheAsnPro 2370
Db 39475 -----CCTCCCTCAGCCAGGAGCTTTCGCGAGGAGCTTCCCTCC 39431
QY 2371 LeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSer 2390
Db 39430 CCCCCTCAGCCAGGAGCTTTCGAGGAGCTTCCCTCCCTCCCTCAGCCAGGAGCTTGTG 39371
QY 2391 AspHisThrLeu-----ThrSerProGly-----GlyGlyGlyLeuAla 2403
Db 39370 CAGGCCACCTTCCCTCTCCCTCAGCCAGGAGCTTTCGCGAGGAGCTTCCCTCC 39311
QY 2404 LysValSerGlyArgProSerArgLysAlaLysSerProAla--ProGlyLeuAla 2423
Db 39310 CCCCCTCAGCCAGGAGCTTCGCGAGGAGCTTCCCTCCCTCCCTCAACCCAGGAGCTTCC 39251
QY 2423 erGlyAspArgPro-----ProSerValSerSerValHisSerGluGlyAspC 2439
Db 39250 CAGGCCACCTTCCCTCCCTCCCTCAGCCAGGAGCTTTCGAGGAGCTTCTCCCTCCCTCC 39191
QY 2439 ysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlyS 2459
Db 39190 TCAACCCAGGAGC-----TCTCAGGCC 39167
QY 2459 erThrProPheProTyAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerP 2479

Db	39166	ACCTCCCTCCCTCCAGCCCA-----GGGACCTCCGACGACCACTCTCCCTCCC	39116	DB:	10	Gaps:	28
Qy	2479	roProProGlyLeuProAlaGlySerGlyProLeu-----AlaGlyProHisHia	2497	US-09-522-753-5 (1-2517) x MMU22016 (1-2036)			
Db	39115	CCCTGAGCCAGGACCTCGCAGGACCACTCTCCCTCTTTAGCCAGGACCTCGCAGG	39056				
Qy	2497	la 2497					
Db	39055	CC 39054					
RESULT 57							
MMU22016							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
TITLE							
AUTHORS							
JOURNAL							
MEDLINE							
PUBMED							
REFERENCE							
TITLE							
JOURNAL							
FEATURES							
source							
gene							
CDS							
polyA_site							
ORIGIN							
Alignment Scores:							
Pred. No.:							
Score:							
Percent Similarity:							
Best-Local Similarity:							
Query Match:							


```
/protein_id="CAC13097.1"
/db_xref="GI:10803169"
/db_xref="GOA:Q9EWY1"
/db_xref="SPTREMBL:Q9EWY1"
/translation="MTAPDSDGLAEQLLRVHRIQKRHLHRLDGLITPAQSRLL
RTLAHWSPPRMADLAERLEVPRAVTLVDGLEAGKGVKVPDPSANRVRIRLELTD
GRAALRLRGARSAEILAPLTDEQRVLGGLDLIDGDTGEGRRIG"
2133..2447
misc_feature
/ gene="SC01147"
/ note="Pfam match to entry PF01047 MarR, MarR family,
score 92.80, E-value 6.9e-24"
complement(2511..3269)
gene
/ gene="SC01146"
/ note="synonym: 2SCG38.39c"
complement(2511..3269)
CDS
/ gene="SC01146"
/ note="2SCG38.39c, possible lipoprotein, len: 252 aa;
similar to TR:Q9RK82 (EMBL:ALJ12662) Streptomyces
coalicolor hypothetical 26.4 kDa protein Scrl1.08, 250 aa;
fasta scores: opt: 645 z-score: 660.7 E(): 2.4e-29; 42.1%
identity in 247 aa overlap. Contains correctly situated
match to Prosite entry PS00013 Prokaryotic membrane
lipoprotein lipid attachment site and possible N-terminal
region signal peptide sequence"
/codon_start=1
/transl_table=1
/product="putative lipoprotein"
/protein_id="CAC13098.1"
/db_xref="GI:10803170"
/db_xref="SPTREMBL:Q9EWY0"
/translation="MARGGGRRTAVAAALGTLTAGTALSACGPGSGDSDRAGBSP
GEARSSASADPTRIQGVDRLOTRIPGDSRQVAVYVGDGSDSADSTVVLTKQGSAM
HRDRWAGHNGTGTDDHENDKRSVGVGTLSDAGGVLPDPPEPAGTGLPYTRGAS
IAAPRWAKWQHDPTVYIAIDYNRVKVTPNDPTPEGTSGKGSILWLHDHSGTSA
CVSLPKPAMEYLLARTLPDQHPVVMGDRHLKA"
complement(3186..3218)
misc_feature
/ gene="SC01146"
/ note="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
3498..5375
gene
/ gene="SC01147"
/ note="synonyms: 2SCG38.40, SCG8A.01"
3498..5375
CDS
/ gene="SC01147"
/ note="2SCG38.40, probable ABC transporter transmembrane
protein (fragment), len: >35 aa
SCG8A.01, partial CDS, probable ABC transporter
transmembrane subunit, len: >623aa; similar to many eg.
TR:O87312 (EMBL:AF027770) PxtA iron sequestration protein
from Mycobacterium smegmatis (574 aa) fasta scores; opt:
794, z-score: 866.3, E(): 0, 35.2% identity in 514 aa
overlap. Contains Pfam match to entry PF00664
ABC membrane, ABC transporter transmembrane region, Pfam
match to entry PF00005 ABC tran, ABC transporters family
Prosite match to PS00211 ABC transporters family
signature. Also contains possible membrane-spanning
hydrophobic regions."
/codon_start=1
/transl_table=1
/product="putative ABC transporter transmembrane subunit"
/protein_id="CAD55163.1"
/db_xref="GI:24413741"
/db_xref="GOA:Q8CK38"
/db_xref="SPTREMBL:Q8CK38"
/translation="MQIQDLPYDPGVPDARSQPRFLWMLFRNLQGLQSLAWGLLH
FASVALPCVGLAIQAVDRSGRLAAGVVMVLCVGNALGDTLHRAAVTNWITA
AARVQQLARKTALLSALTRVAAGVAVSTGDEKIGVFVEAVSRFTAAALTIVL
VCVALVYQPALGVVLAVLAVLPILPRATRADVRDKRAGRATLADSTVAG
LRLRGIGBELFLDRYDASQVRHAAVRSARMSLSIAIOVLPLCLLIIAVVHWGV
HLAREGRIDVGLVTVVSYVMILTYPLRHFEETAMAYFSRPSAKRAARVLSERTTD
TEGSRRAEPTGLDYPATGLLAPAGFTAVVCGDPAAGRLAERLGGHPSRGTSAL
LGVPLDDELPLKSAFTAVLDQKPDVLSGLTRELIDVPASGAVGAEADLASAQCQGV
LTLAQSGVGADDMARDITERGSRISLGGQRQLALARSLLITDPEALVLDPEPTSAVDS
```

Db	282398	AGGTACAGAGCGGTACGACGCGCCACACACCGCGCGCGGTCCACACCGCCACGCTAC	2823398
Qy	209	-----GluProGluLysProValSerProProIoi	219
Db	282338	GTCGTCGACGTCACGGTCGCGTACACGACGCGCCACACGCTCCAGACCTCCACGCGGT	282279
Qy	219	leGluSerLys-HisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAla	238
Db	282278	CGGTCGACGCGCCACCGTCCG-----CGGCGGTCGCC	282246
Qy	239	GluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsn	258
Db	282245	GAG-----CTTCGCGC-----GGCGCGTCGCGGTACAGGCGTCCCGTA	282204
Qy	259	GlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLys	278
Db	282203	GCGCGCGGACATCAGACAGCAG-----GAAGCGGAGCGGTGCACCGACCGCGGAA	282150
Qy	279	LysLeuIleLeuTyrPheLysArgAsnHisAlaArgLysGlnTrpLysGlnLysPhe	298
Db	282149	CCGTCGCGTTCAGCAGCACACCCCGACACGCTCCGCGCACCGGTACAGCGC---	282093
Qy	299	CysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsn	318
Db	282092	-----TCG	282090
Qy	319	AsnProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPhePro	338
Db	282089	CACCGCGCGCGCGGCGCTCGCGGCAACAGCGC-----	282054
Qy	339	GluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGly	358
Db	282053	-----GCCTCGCGCGCACCACTGCCGTACCG-----CGGCGCGCGCGGAGTCCGAC	282003
Qy	359	SerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGly	378
Db	282002	GAG-----GCCGGGGAGCGCGGCGAAGCTACAC-----GTCCGCGCGCGG	281961
Qy	379	LeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProPro---	397
Db	281960	CAGCGCGCGCGCGCGGAGAACAGCTCCCTCGCGCCAACTCGCTCCCGCGGACCGCGC	281901
Qy	398	-----MetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsn	411
Db	281900	CACGGCGCACTCGCGCTCCGCTCCACCGCTCGACGCCACAG-----	281856
Qy	412	MetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMet	431
Db	281855	-----CGCCCCCGCACCGCTC-----	281841
Qy	432	TrpSerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPhe	451
Db	281840	-----GCACACACCG-----	281829
Qy	452	GlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyr	471
Db	281829	-----	281829
Qy	472	LeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgGly	491
Db	281828	-----CATCGCGCGCACCGCGCGCGCG	281805
Qy	492	LysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	509
Db	281804	CGCCACAGACAGCGGTGAGCTTACCAAGTCCCGCGCGCAGGTGGCGCAGCATCGCNA	281745
Qy	510	GlnProMetProArgSerSerGlnGluGluLysAsp---GluLysGluLysGluLysGlu	528
Db	281744	CGGCGCGTCCCGTCAACCGACCCCGTCCCGCGCATCCGGAATCCCTGCGCCCGCAGCAC	281685
Qy	529	AlaGluLysGluGluGluLysProGlu-----	537

281684	DB	GCCACGCGTGTGAGCAGGCGCCCGCCGGCA	CGTGTGCTGTACGCGGGGCACACGTC	CCAG	281622
538	QY	-----ValGluAsnAspLysGlu	543	-----	
281624	DB	CGTCACCTCCACGGCAGCCGCCCGCTCGCGC	CACGGCGTGTGCAGGGGCTGGAGACC	281565	
544	QY	AspLeuLeuLysGluLysThrAspAsp-Thr	-----SerGlyGluAspAspAspG	560	
281564	DB	ACCGAGTCCGCGAGCGTGC	CCGCGAACAAGTTGACGTG	CACGGCGCAGCAGCGTCTCCG	281505
560	QY	uLysGluAlaVal	-----AlaSerLysGlyArgLysThrAla	AsnSerGlnGlyArg	578
281504	DB	CGCGCCCGCCGACGAGCCGACACCGCAGC	AGCCGCTCGAGTTCGGATCGCGCG	281445	
578	QY	gLySGlyArgIleThrArgSerMetAlaAsn	GluAlaAsnSerGluGluAlaIleThrPr	598	
281444	DB	CGCTCGCGCAGCAGTGC	CGCGTCTCCGGCTCGCCAGTATC	CACGGCGCGGACCC	281385
598	QY	oGln	-----GlnSerAlaGluLeuAlaSerMet	GluLeuAs	610
281384	DB	CGCGGTTCATCAGGTTGACCA	CCGCTTGGAGGCGAAGCGGAGAGTG	TCGTCGTCAGAACG	281325
610	QY	nGluSerSerArgTrpThrGluGluMet	GluThrAlaLysLysGlyLeuLeuGluHi	630	
281324	DB	CACGGGACAGGATGG	-----CGCGCAGGCTCCCCA	281292	
630	QY	sGlyArgAsn	-----TrpSerAlaIleAlaAr	639	
281291	DB	CGGCGCGGTTCACGCGAGCTTC	CGGATTCGGGTGCGTCA	CGCGGCGTGTGCCG	281232
639	QY	gMetValGlySerLysThrValSer	-----GlnCysLysAsnPh	652	
281231	DB	ACGGGTTCGCGCGCGCCCACTTCCCT	CCCGAGCCCGGCTGATAAATCGAGGA	ACA	281172
652	QY	eTyrPheAsnTrpLysLysArgGln	AsnLeuAspGluIleLeuGlnGln	-----	668
281171	DB	T	-----GACAGACTTC	CGCGCGCTCCCGCGTC	281142
669	QY	HisLysLeuLysMetGluLysGluArg	AsnAlaArgArgLysLysLysLys	-----LysAlaP	687
281141	DB	TGGCCACACCTTCGCGCCCTGT	CGCGCTCAC	-----CGCGTCCCGCGGTCGCGCC	281085
687	QY	roAlaAlaAspGluGluAlaAlaPhePro	ProValValGluAspGluGluMetGluA	707	
281084	DB	CGCGCACCGCG	-----CAGGCACGGCACGACCGC	GAGAGTTCGTGGGCC	281031
707	QY	LaSerGlyValSerGlyAsnGluGluMet	ValGluGlu	-----	720
281030	DB	TGCGGGCGTC	-----GACCCGACGATCATCGAGGAGAT	GCCTACGTCACCGCGC	280980
720	QY	-----	-----	-----	720
280979	DB	ACAACTTCGTCGGGAGCGCTCGAC	GGCTACCGGCAGCCCCGTGTGCATCTCT	CACCCCGC	280920
721	QY	AlaGluAlaLeuHis	-----	-----	725
280919	DB	CGCGCGCGGAAGCGCTCCACCT	CGCCACAGCCCGGCTGCTGGCGCAGGGCT	ACTCCTCA	280860
726	QY	-----AlaSerGlyAsnGluValPro	-----ArgGlyGluCysSerGlyPro	AlaLys	740
280859	DB	AGGTGTACACTGTCTACCGGCC	CCGCGCGTGCACCACTTCGTGGCTGGCGGAAG	280800	
741	QY	ThrValAsnAsnSerSerAspThr	GluSerIleProSerProHisThr	GluAlaLys	760
280759	DB	ACCTCGACGACGAGACATGAAG	CGCGAGTTCACCGCAGCTGCACAAGAC	CCCGCTCT	280740
761	QY	AspThrGlyGlnAsnGlyProLysPro	ProAlaThrLeuGlyValaAsp	-----	776
280739	DB	TCGCCGACGGCTACATCCCG	GAGAGTCCGGCCACAGCGCGGCTCC	ACCGTGGACCTCA	280680
777	QY	-----GlyProProGlyProPro	-----ThrPro	-----	785
280679	DB	CCCTCGTACGAGCTCCGCGCA	AACCAACCGCGCTGACCA	CCCGCGGACCCCTGGTGC	280620

```
Qy 786 -----ProArgArgThrSerArgAlaProIleGluPro----- 796
Db 280619 CCTGTTCCGCCCGAAGGGGAGCGTTTCCCGGACAACTCGTCACATGGCGACCGGAT 280560
Qy 797 -----ThrProAlaSerGluAlaThrGlyAla 805
Db 280559 ACGACTGTTCCGACACCGCGTTCGACACCCCTCGACCCCGGATCGAGGGCGTCCAGCGTG 280500
Qy 806 ProThr----- 807
Db 280499 CCAACCGGCTGCTCTCAAGACACCCCTGGAGGAGGGCGGCTGGTGAACCTCCCGGAGG 280440
Qy 808 -----ProProProAlaProProSerProSerAlaProProProValValProLys 824
Db 280439 AATGGTGGCACTTCCACCTACAGGCCCGAGCGCTTACCGGACACCTACTTCGACTTCCCGG 280380
Qy 825 GluGluLysGluGluGluThrAlaAla----- 834
Db 280379 TTTCGCGGAAGTCTCTCCGCGCGCGCTGACCGGCTGCGGAGCGGCTCCACTCCGCC 280320
Qy 835 -----ProPro-ValGluGluGluGluGluGluProProAlaAla----- 848
Db 280319 ATCGGATACAGTCCGCGCGTGTCCGAAATTGAAAACTTCGTTCCCACTCCGCGCTCGG 280260
Qy 848 ----- 848
Db 280259 CTCCCACTGTTGAGCTGCGGATCGCCTACGGAGAGGGGCTCTCCGCTGGCCCCGCGAC 280200
Qy 848 ----- 848
Db 280199 CTGTCCCGGTGCGGCACCGTGGCTTACCGCAACCCGCTGCGCGTGGCGCTCCCTCA 280140
Qy 848 ----- 848
Db 280139 GCCGTGTACACACGAGGCGACCGCTGCTGCTCATCACCGCACCGTGACCCCGCG 280080
Qy 848 ----- 848
Db 280079 GCGCGGGGCGTGCCCTGCCCGCGCTACGTTCGACGACCGGGAGGACTGGCGGCGAGC 280020
Qy 849 -----GluGluLeuAlaValAspThrGlyLysAlaGluGluProVa 862
Db 280019 CGTGTCCCGAGCTCAAGGAGGAGACCGGCTCGACGC-GGCGACCGCGACGTCGCGC 279961
Qy 862 LysSerGluCysThrGluGluAlaGluGluProAlaLysGlyLysAspAlaGluAl 882
Db 279960 TCGCCGAGCG---CATGAGCTCGCCCGACGGCCACCTGCTGCTTTCGGGCTGCTCCCG 279904
Qy 882 aAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysGluGlyGlySerGly-- 901
Db 279903 AGCGCGCGCGGAGGAGTGCCTGCTCCGCGCGCACGAGGACCG-GGGGCTGGCAC 279845
Qy 902 -----ArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAs 916
Db 279844 CTCTCTGCGAGTTCGGAAGAGCTCGCTTCCCGCTCCACAGGCGCTACGCGCCTTC 279785
Qy 916 pSerSerAlaThrCysSerAlaAspGluAlaValAspAlaGluGly-----GlyAs 933
Db 279784 TTCAGGCGCGCTACATCTGAGCGCGAGCTCCCGGAGCCACGAGCGGACCGGGCAG 279725
Qy 933 pLysAsnArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAl 953
Db 279724 GACGCTCGCTCAGACCGTCTCTCGCCCTCCCGCT-----CCACGACG 279683
Qy 953 aAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaI 973
Db 279682 ACCGCGCGCCCT-----CCCAGCGGACACGTACCGTTCTGA-----TCTCCGGT 279638
Qy 973 eProProIleGlnValThrLysValHisGluProProArgGluAspAlaAlaProThrLy 993
Db 279637 TCACACCGCGCTGCGCGCTCCGCGACGACACCGTCCACCGCGCGTCCGCGCCCGCGCG 279578
```

```
Qy 993 s----ProAlaProProAlaPro-----ProProProGlnAsnLeuGlnProGluSe 1009
Db 279577 GGGCGCCACACCTCGAGCTCGGCGCGCGCTCGGCCCA-----GGACCGGAATC 279527
Qy 1009 rAspAlaPro-----GlnGlnProGlySerSerProArgGlyLysSerArgSerProAl 1027
Db 279526 ACCGCTCCCGCGCGCGAAACACCGGAATCCGCCGAAGCGCGTCCAGCACCTGTC 279467
Qy 1027 aPro-----ProAlaAspLysGluAlaAlaPheAlaAlaGluAl 1039
Db 279466 GCGGTCCTTCGTACGCGCGCTCCGCTCGCCACGCTGTCACAGCGCGCGCGGACACCGC 279407
Qy 1039 aGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValProPr 1059
Db 279406 ACCGCAC---GCGGTCGCGCGCGCTCGAGCACCGGGGCGCACGACAGTCCCGC 279350
Qy 1059 oArgGluValLysLeuAspProHisAlaProAspProSerAlaPheSerTrpAlaPr 1079
Db 279349 AGCAGAAACCGCTCTCGCAGTCCCGCAGCATCCGCTCTCCGGCGTCCGACCCACAGG 279290
Qy 1079 oProGlyHisProLeuGlyLeuHisAspThrAlaArgProValLeuPro-Argp 1099
Db 279289 GGACGCGCAC-----AGGAGACACCCGTACGCGCGCGCAGTCCGCGCGTCC 279242
Qy 1099 roProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuG 1119
Db 279241 ACGAAGTACGCGACGCGCGCGCTC-----GACGAGCGCACCGCGCGTGTCTCC 279188
Qy 1119 luArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTrpSerG 1139
Db 279187 AGCAC----- 279183
Qy 1139 luHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProL 1159
Db 279182 -----CTCGTCCCGAATCCACCGCTCCCGTCCGCGC----- 279150
Qy 1159 ysLysLeuAlaProPheSerGlyValLysGlnGlnLeuSerProArgGlyGlnAlaG 1179
Db 279149 -----GGCGCGCGCGCGGTGGTGGGAAACAGCGCAGGTACCGCGCGCAGTGC 279098
Qy 1179 lyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGly---T 1198
Db 279097 AACACCGCGAG-----ATACAGCTCCGCGGAAGAA 279068
Qy 1198 hrAlaLeuGlySerValProGlySerIleThrLysGlyIleProSerThrArgValp 1218
Db 279067 CCCCCCTCGAGCGG---CCCGCATCCGCGCGCGGAGACGCGCACCGCGCGAGCC 279010
Qy 1218 roSerAspSerAlaIleThrTyArgGlySerIleThrHisGlyThrProAlaAspValL 1238
Db 279009 CCAGCACCGTGGCAGCGCGCGCAACCC-----GGCACCGCGTGGCGCTT- 278960
Qy 1238 euTyLysGlyThrIleThrArgIleIleGlyLeuAspSerProSerArgLeuAspArgG 1258
Db 278960 ----- 278960
Qy 1258 lyArgGluAspSerLeuPro-----LysGlyHisValIleTyArgGlyL 1273
Db 278959 --GCCCGGACAGGTGCGCGCTAGCGTGCAGCGCGCGCGCCACCGGACGAGAGGA 278902
Qy 1273 ysLysGlyHisValLeuSerTyArgGlyMetSerValThrGlnCysSerLysGluA 1293
Db 278901 CGAAGCGC-----CGCTCGT 278887
Qy 1293 spGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyArgPM 1313
Db 278886 CGGGGCGCAGTGCAGCGCCCTCGTATCCGCGCGCGCCAT-GCAGCGCGTAGACG 278828
Qy 1313 etMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyA 1333
Db 278827 TTGTGCGCTCCCTGTG-----GTCACCGCGCGCGCTCCAGCGATCGCGGCC 278777
Qy 1333 rgAlaIleProProGluArgHisSerProHisLysLeuLysGluGlnHisIleArgG 1353
```

```
Db 278776 GA-----CGGGGACGCTCGACTCCCC-----GAAGGCGCGAACGAG 278738
QY 1353 lYSerIleThrGlnGlyIleProArgSer-----TyrValG 1365
Db 278737 GTGGGTCGTCCAGGTGCGTCCAGAACCGGTTAAACCCCTGCCGAGCGCTCTCTCGTAG 278678
QY 1365 luAlaGlnGluAspTyrLeuArgArgGluAlaAlaLysLeuLeuLysArgGluGlyIleThrPro- 1384
Db 278677 AGCCCGCCCAACCACTCGCGCGGACACCGCGTGGTGAAGTCCGGGAAGACCACTCCGCC 278618
QY 1385 -----ProProProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuG 1403
Db 278617 GGACGCCCCACCCCGCGGACGACCGT----- 278592
QY 1403 lyProLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyA 1423
Db 278591 --CCCGCCGCATCCCGCACGAACAC-----GTCCCGGCGCCACCG 278552
QY 1423 rgSerIleHisGluIle-----ProArgGluGluLeu-----ArgHisThrP 1437
Db 278551 CCCTCGTACAGAGTTGCGGGCGCGGCCGGAACCGCGGCTCCACGGCCGACACACGAG 278492
QY 1437 roGluLeuProLeuAlaProArgProLeuLysGlu---GlySerIleThrGlnGlyThrP 1456
Db 278491 CGGAC---CCCTCCCTCGCGAGCTCTCGCGCAGTACGGGCGAGTTCGGGAA-----C 278441
QY 1456 roLeuLysTyrAsp-----ThrGlyAlaSerThrThrGlySerLysLysHisA 1472
Db 278440 CGCTCTCTCGTACGCGGTGAACACTCGCGGCGAGTGCCTGCCGATGCCAGATGCACC 278381
QY 1472 spValArgSerLeuIleGlySerPro---GlyArgThrPheProProValHisPro----- 1489
Db 278380 CGCTCCAGGGCAGATCCC-GCTCTGTGTGCGCGCCACGACCGCGCTCAGATCCTGCTC 278322
QY 1490 -----LeuAspValMetAlaAspAlaArgA 1498
Db 278321 GTCCGCCGAATCCACAGCGGTGATGATACCGAGCGCCACCGCGGCGAGCGCGGG 278262
QY 1498 laLeuGluArgAlaCysTyrGluGluSerLeuLysSerArg-----ProG 1513
Db 278261 CCCTCCGCTACGAGCGCCAGGCGGAGCAGCAGCGCGGGGTACCCACCGTCAACCCA 278202
QY 1513 lyThrAlaSerSerGlyGlySerIleAlaArgGlyAlaProVal-IleValProGlu 1532
Db 278201 GCACGCGAGCGGACCGCGTCCATCCGAGCAGCTCCGCGCGCGGTGTCGCCCGGA 278142
QY 1533 LeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGly 1552
Db 278141 CCGCGGCC-CTTCCTCACCTCCCGCAACG-----CCACCGTGCCTGCC 278098
QY 1553 HisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGly 1572
Db 278097 ACGTGCTGTCGTGGAAGACAGTGCCTCCGCGCTCCGCCACCAACCACTGCACCGGCA 278038
QY 1573 SerLeuSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIle 1592
Db 278037 TCGTGATGACACGCGGACGCGTCTCTCGCGCGCGCGGAGGCGAGCGGACCGG 277978
QY 1593 AlaLysSer-ProHisSerThrValProGluHisHisProHis-----ProIleSerPr 1610
Db 277977 TGTTCACAGCGGTACGTGCTCTCGCGCAGCGGGGCGCGCGAGCGCGCGAGCC 277918
QY 1610 oTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIle--Pro 1629
Db 277917 CGAAGAA-----CCGGGCGTC-----CGCGCCACTCCGCC 277885
QY 1630 LeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaIleThr 1649
Db 277884 GCTGCTCCAGCGGGGAGACCGCCCGGCTCCACCCAGCGGGCGGCGAGTCACTC 277825
QY 1650 TyrLeu-----ProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrPro 1666
```



```
Qy 1972 -----AlaSerSerProSerLysGlySerGluProArgProLeuValProValSer 1989
Db 276858 GGACGAACACCGCTCTGTAGTCGAGAGACCAACGA---ACCGG-CCCGCGTCACCT 276803
Qy 1990 GlyHisAlaThrIleAlaArgThrProAlaLys-----AsnLeuAlaPro----- 2004
Db 276802 GGGCAGAGCTCGCGCGCAGTCGCTCCCTCCCTCGCGCGCAACTCGCGCGCTCGCGGTAC 276743
Qy 2005 -----HisHisAla 2007
Db 276742 GCGCGCGGACCGGTCAGCGGTCACCTCCCAACATCCCGCAGCGCGTGTGCGGTGC 276683
Qy 2008 SerProAspProPro-----AlaProProAla-----Ser 2017
Db 276682 TCGCACCGCGCGCGTGGCGCGCTCTGGACCTCTCGCGCGCGACTTCGGCGCCCGCA 276623
Qy 2018 AlaSerAspProHisArgGluThrGln-SerLysProPheSerIleGlnGluLeuG1 2037
Db 276622 GCGTCTCGACCGCTTCCAGCAGGTGCAACCGCTGCTCTGTTCCACCGTCGACGGCTACC 276563
Qy 2037 uLeuArgSerLeuGlyTyHis-----GlySerSe 2047
Db 276562 GCTACGGGGGCAAGAAACACACCGCGCGGACCGCTCGCGAACTCGCGCGGAAGTGC 276503
Qy 2047 rTySerProGluGly---ValGluProValSerProValSerProSerLeuThrHi 2066
Db 276502 CCACCTCCGCGGCTGATCCACATCCCGCTCTCGGCACCGCGCGCGCGCGC 276443
Qy 2066 sAspLysGlyLeuProLysHisLeuGluLeuAspLysSerHisLeuGluGlyLe 2086
Db 276442 TGGACTGGGAGACCT---CACCGCGCGGACCGCGGTGTACGAGCAGGTGCGCT 276386
Qy 2086 uArgProLysGlnProGlyProValLysLeuGlyGluAlaAlaHisLeuProHisLe 2106
Db 276385 TCGACCAACCCCTGTGGTGTCTACTCTCCGGCACACCGCGCTCCCAAGGCCAT-- 276328
Qy 2106 uArgProLeuGluSerGlnProSerSerProLeuLeuGlnThrAla---ProG1 2125
Db 276327 -CGTCCAGTCCCGGCGGATCTCTCGAGCACCT-----CAAGCAGTCCGCGCTGC 276275
Qy 2125 yValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrG1 2145
Db 276274 ACTCGACCTCGGCGCGGACCGCTTCTCTGTATACCTCGACCGGCTGGATGATGT 276215
Qy 2145 nAspTyThr-----ArgHisHis-----ProGlnGlnLeuSerAl 2157
Db 276214 GGAATTCCTCGTCTCGGCTGTCTACCGGCACACGATCGTCTCTACGAGC-GCAGC 276156
Qy 2157 aProLeu---ProAlaProLeuTySerPheProGlyAlaSerCysProValLeuAspLe 2176
Db 276155 CCGGCTTCCCGCACCGACCGCCAGTGGCGCATCGGAGACGACCGCGGCCACCCCTT 276096
Qy 2176 uArgArgProProSerAspLeuTyLeuProProAsp-HisGlyAlaPro---AlaA 2195
Db 276095 TTGGGACCT-----CGCGCGCTACGTCATCGTGGCTCGCGAAGGCC 276054
Qy 2195 rgGlySerProHisSerGluGlyGly-----LysArgS 2206
Db 276053 GGCGTCCACCGCGCGGACCTGAGCTCTCCCGATCCAGTGTGTCGCACACCGCGC 275994
Qy 2206 erProGluProAsnLysThr-SerValLeuGlyGlyGlyGluAspGlyIleGluProValS 2226
Db 275993 TCCCGCTCCCGCGCGGCTCCGCTGGCTGTCGACGAGTTCGCGCGCGGCGTGC 275934
Qy 2226 erProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyProLeuLeuT 2246
Db 275933 GACCTGTGGATCGCTCCGTGAGCGCGCGGACCGACGT-----GTGCTCC 275889
Qy 2246 yArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGly-----A 2264
Db 275888 TGCTTCGCGGCGCGCTCCCGACCTCCCGTCCACATCGCGGAAGTCCAGGCCCCCGCGC 275829
```

```
2264 snThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValL 2284
275828 CTGGCACCGACCTCCA-----GTCTGTG 275805
Qy 2284 ySerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluT 2304
275804 GACCCGAGCGCAACCCCTTGACCGAGGTCGGCGAGCTGTGTCACCAACCCGATG 275745
Qy 2304 yrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyGlyL 2324
275744 CCCTCGATGCGGATCCGCTTCTGGAACG-ACCCTGACGGCGCGCTTACCACGACGCTA 275686
Qy 2324 euMetThrTyArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluA 2344
275685 CTTTCGACA-----CCTACCCCGCGCTGTGCGGCGCGCGACTGGA----- 275645
Qy 2344 laIleIleArgLysAlaLeuMetGlyLysTyArgGlnTrpGluGluSerProProLeuS 2364
275644 -----TCACCTTCACCTCCCGCGG 275626
Qy 2364 erAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleT 2384
275625 CTCCG-----TCGTATCCACGGCGGCTCCGACTCCACCTCAACCG 275584
Qy 2384 hrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaL 2404
275583 CCAAGGCGTCCGATGGCTCCGCGGACATCTACGAGGCGGTGCGAGCGGTCCCGGAGA- 275525
Qy 2404 ySValSerGlyArgProSerSerArgLysAlaLysSerPro-----AlaProGlyLeuA 2422
275524 -----TCAGGGAATCCCTCGTCATCGGCATCGAACACCGCGCGCTACTCGATGCC 275470
Qy 2422 laSerGly-----AspArgProProSerVal-----SerSerValHisSerG 2436
275469 CTTCTTCGTGCACCTGCGACCCCGGAGCCACCTCGACGACGCGCTCTGACCCGATCAA 275410
Qy 2436 luGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerS 2456
275409 GCGGACCATCCGCGCAACCTCTCCCGCGCCACGCTCCCGACGAGGTCTATCGAGGTCCC 275350
Qy 2456 erAlaGlySerThrProPheProTyArgProLeuIleMetArgLeuGlnAlaGlyValM 2476
275349 CGGCATCCCGCACACCTCCACCGCAAC-----GCATCGAGGT 275311
Qy 2476 etAlaSer-ProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHis 2495
275310 CCGGTCAACGGCTCTCTCCAGGCGACCCGCTGGACAGCGGTCAACCCCGCTCAT 275251
Qy 2496 HisAlaTrpAspGluGluProLysProLeuLeu 2506
275250 CGA-----CAACTCGACCTGCTCCACTTCTA 275224

RESULT 60
AC106029 180591 bp DNA linear PRI 06-AUG-2002
LOCUS Homo sapiens chromosome 15, clone CTD-2538111, complete sequence.
DEFINITION AC106029
ACCESSION AC106029.8 GI:22094423
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 180591)
AUTHORS Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N.,
TITLE Anderson,S., Barna,N., Bastion,V., Boguslavsky,L., Boukhalter,B.,
JOURNAL Brown,A., Camarata,J., Campopiano,A., Chang,J., Charaz,B.,
REFERENCE Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
2 (bases 1 to 180591)
AUTHORS Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
```

Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,
 Lamazars, R., Landers, T., Lechoczy, J., Levine, R., Liu, G.,
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,
 Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (11-JAN-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 180591)
 Birren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
 Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
 Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
 Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
 Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
 McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
 O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
 Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P.,
 Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
 Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
 Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (bases 1 to 180591)
 Birren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
 Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
 Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
 Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
 Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
 McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
 O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
 Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P.,
 Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
 Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
 Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (03-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 5 (bases 1 to 180591)
 Birren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
 Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
 Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
 Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
 Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
 McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
 O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
 Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P.,
 Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
 Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
 Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission

TITLE
 JOURNAL
 COMMENT
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L24001
 Center clone name: 2538_I_11
 ----- Location/Qualifiers
 1. 180591
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="15"
 /map="15"
 /clone="CTD-2538I11"
 /clone_lib="CITD2 Human BAC"
 715..740
 /rpt_family="SUBTEL_sat"
 repeat_region
 746..1009
 /rpt_family="TAR1"
 repeat_region
 1019..1044
 /rpt_family="TGAA)n"
 repeat_region
 1068..1212
 /rpt_family="MER5B"
 repeat_region
 complement(1765..2037)
 /rpt_family="AluJb"
 repeat_region
 2423..2594
 /rpt_family="MIR"
 repeat_region
 complement(3529..3745)
 /rpt_family="AluSg/x"
 repeat_region
 complement(4320..4681)
 /rpt_family="THE1C"
 repeat_region
 5527..5646
 /rpt_family="MIR"
 repeat_region
 complement(5682..5862)
 /rpt_family="AluJo"
 repeat_region
 complement(5897..6201)
 /rpt_family="AluSg"
 repeat_region
 complement(6413..6484)
 /rpt_family="AluS"
 repeat_region
 6485..6606
 /rpt_family="CT-rich"
 repeat_region
 complement(6607..6747)
 /rpt_family="AluJo"
 repeat_region
 complement(6748..7046)
 /rpt_family="AluSx"
 repeat_region
 complement(7047..7183)
 /rpt_family="AluJo"
 repeat_region
 8623..8643
 /rpt_family="(TTG)n"
 repeat_region
 complement(8690..8984)
 /rpt_family="AluSx"
 repeat_region
 complement(9670..9970)
 /rpt_family="AluSx"

```

repeat_region complement(10259..10512)
repeat_region /rpt_family="MER46C"
repeat_region complement(10635..10802)
repeat_region /rpt_family="MIR"
repeat_region complement(11162..11476)
repeat_region /rpt_family="AluSx"
repeat_region 11477..11963
repeat_region /rpt_family="Trigger3 (Golem)"
repeat_region complement(11996..12161)
repeat_region /rpt_family="MIR"
repeat_region 12270..12563
repeat_region /rpt_family="AluY"
repeat_region 12959..12993
repeat_region /rpt_family="GC-rich"
repeat_region 13464..13497
repeat_region /rpt_family="AluSx"
repeat_region 14055..14171
repeat_region /rpt_family="(TG)n"
repeat_region /rpt_family="L2"
repeat_region 14364..14650
repeat_region /rpt_family="AluJo"
repeat_region complement(14920..15228)
repeat_region /rpt_family="AluSx"
repeat_region complement(15232..15518)
repeat_region /rpt_family="AluSx"
repeat_region 15672..15743
repeat_region /rpt_family="(TA)n"
repeat_region complement(15744..16059)
repeat_region /rpt_family="AluSx"
repeat_region complement(16069..16195)
repeat_region /rpt_family="FLAM_C"
repeat_region 16196..16224
repeat_region /rpt_family="AT-rich"
repeat_region complement(16411..16962)
repeat_region /rpt_family="L1MB7"
repeat_region complement(16963..17126)

```

Alignment Scores:

```

Pred. No.: 1.6e-05 Length: 180591
Score: 715.50 Matches: 566
Percent Similarity: 32.70% Conservative: 307
Best Local Similarity: 21.20% Mismatches: 1112
Query Match: 5.41% Indels: 694
DB: 9 Gaps: 115

```

US-09-522-753-5 (1-2517) x AC106029 (1-180591)

```

Qy 14 AlaThrGluProArgTyrProHisSerLeuSerTyr-ProValGlnIleAlaArgTh 33
Db 6885 GCCAGCTGGCCAGACAGCAGCCAGCCACTCAGCCAGCCAAAGCCAGCCAGCCAGCC 68944
Qy 33 rHisThrAspValGlyLeuLeuGluTyrGlnHisSerArgAspTyrAlaSerHisLe 53
Db 68945 ACCAGT-----CAGCCAGCCACTCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 68989
Qy 53 uSerProGlySer---lleGlnPro---GlnArgArgProSerLeuLeuSerGl 71
Db 68990 CCAGCCAGCCAGCCAGCCCTTCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69038
Qy 71 uPheGlnProGlyAsnGluArgSerGlnGluLeuHis---LeuArgProGluSerHis 90
Db 69039 ----CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69092
Qy 90 rTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProArgLe 110
Db 69093 -----CCGCCGCGC-----CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69121
Qy 110 uGluLeuLeuPro---AspProLeuLeuArgProSer-----ProLeuLeuAl 125
Db 69122 GCCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 69181
Qy 125 aThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLys-- 144
Db 69182 GGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69241

```

```

Qy 145 -LeuGluProValSerProProSerProHisThrAspPro----- 158
Db 69242 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 69301
Qy 159 -----GluLeuGluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMe 176
Db 69302 CCAAGGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69341
Qy 176 tAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLeY 196
Db 69342 -----ACACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 69394
Qy 196 sGlnGlnGlnLeuGluGluAlaAlaLysProProGluProGluLysProValSerPr 216
Db 69395 CACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 69454
Qy 216 oProProIleGluSerLysHisArgSerLeuValGlnIleLeuTyrAspGluAsnArgLy 236
Db 69455 GCCAAGC---CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 69493
Qy 236 sLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLe 256
Db 69494 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 69539
Qy 256 uTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMe 276
Db 69540 ----AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 69551
Qy 276 tArgLysLysLeuLeuLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGl 296
Db 69552 -----CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 69582
Qy 296 nLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuLysLysValGluArgIl 316
Db 69583 A-----GCCCGCCA 69591
Qy 316 eGluAsnAsnProArgArgAlaLysGlySerLysValArgGluTyrTyrGluLysGl 336
Db 69592 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 69628
Qy 336 nPheProGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGl 356
Db 69629 ----CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 69684
Qy 356 nArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIl 376
Db 69685 CAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 69727
Qy 376 eAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIlePr 396
Db 69728 -----CTAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69780
Qy 396 oProMetLeuTyrAspAlaAspGlnGlnArgLysPheIleAsnMetAsnGlyLeuMe 416
Db 69781 AAGCCTGTCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69819
Qy 416 tAlaAspProMetLysValTyrLys-----AspArgGlnValMetAsnMetTr 432
Db 69820 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69879
Qy 432 pSerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGl 452
Db 69880 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 69938
Qy 452 YLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrLe 472
Db 69939 CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 69951
Qy 472 uThrLysLysAsnGluAsnTyrLysSerLeuValArgSerTyrArgArgArgGlyLys 492
Db 69952 -----CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 69959

```


QY	1843	yGlyGlySerSerSerArgProAlaSerHisSerHis-----AlaHisGlnHisSe	1861
Db	73668	AGCCAAAGCCAGCA-AGCCAGCCAGCCAAAGCCAGCCAGCCAAAGCCAGCCAGCCAA	73726
QY	1861	rProIleSerProArgThrGlnAspAlaLeuGlnArgProSer-----ValLeuHi	1879
Db	73727	GCCAGCAACCCAGAGAGCCAGCCAAATCCAGCCTAGCAGCCAGCCAAAGCCAGCCAGTCA	73786
QY	1879	sAsnThrGlyMetLysGlyLeuThrAlaValGluProSerLysProThrValLeuAr	1899
Db	73787	TGCAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	73846
QY	1899	gSerThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCy	1919
Db	73847	AGCCAGCCAAATCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	73895
QY	1919	sProLeuGlyGlyThrLeuAspGlyValTyProThrLeuMetGluProValLeuLeuPr	1939
Db	73896	-CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	73948
QY	1939	oLysGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPh	1959
Db	73949	CAGACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	74008
QY	1959	eLeuAlaLysProProAlaArg-----SerGlyLeuGluProAlaSerSerProSerLy	1977
Db	74009	CCAGTCACCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74067
QY	1977	sGlySerGluPro-----	1981
Db	74068	GGCAAGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	74127
QY	1982	-----ArgProLeuValPro-ProVal-----SerGlyHisAlaThrIleAlaArgThrP	1998
Db	74128	AGTCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74187
QY	1998	roAlaLysAsnLeuAlaProHisAlaSerProAspProProAlaProProAlaSer-	2017
Db	74188	CAGCCAAAGTCA-CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74244
QY	2018	-AlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuG	2037
Db	74245	AAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74282
QY	2037	luLeuArgSerLeuGlyTyHisGlySer-----	2048
Db	74283	-----CACTCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74322
QY	2048	yrSerProGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspL	2068
Db	74323	AAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74369
QY	2068	ysGlyLeuProLysHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArgP	2088
Db	74370	-----CAGC	74373
QY	2088	roLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgP	2108
Db	74374	CAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74433
QY	2108	roLeuProGluSerGlnProSerSerProLeuLeuGlnThrAlaProGlyValLysG	2128
Db	74434	AAAGCCCGCCAAACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAA	74492
QY	2128	lyHisGlnArgValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrt	2148
Db	74493	-----GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74529
QY	2148	hrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrsPheProG	2168
Db	74530	AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74576
QY	2168	lyAlaSerCysProValLeuAspLeuArgProProSerAspLeuTyrtLeuProProp	2188
Db	74577	-----AAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74595
QY	2188	roAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProG	2208
Db	74596	GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74655
QY	2208	luProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProp	2228
Db	74656	AGCCAAACCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74696
QY	2228	roGluGlyMetThrGluProGlyHisSerArgSerAlaValTyProLeuLeuTyrtArgA	2248
Db	74697	-----GCTAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	74724
QY	2248	spGlyGluGlnThrGluProSerArgMetGlySer-LysSerProGlyAsnThrSerGln	2267
Db	74725	CAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74784
QY	2268	ProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLys	2287
Db	74785	CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74843
QY	2288	GlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrtAsnIleSer	2307
Db	74844	CAACCA-----AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74873
QY	2308	GlnProGlyThrGluIlePhe-----AsnMetProAlaIleThrGlyThrGlyLeuMet	2325
Db	74874	CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74933
QY	2326	ThrTyrtArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIle	2345
Db	74934	CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74957
QY	2346	IleArgLysAlaLeuMetClyLysTyrtAspGlnTyrGluGluSerProProLeuSerAla	2365
Db	74957	-----	74957
QY	2366	AsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAla	2385
Db	74958	-----GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74990
QY	2386	AlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLysVal	2405
Db	74991	AGC-----CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	75044
QY	2406	SerGlyArgProSerSerArgLysAlaLysSer-----ProAlaProGlyLeuAlaSer	2423
Db	75045	CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	75104
QY	2424	GlyAspArgProProSerValSerValHisSerGlu-----GlyAspCysAsnArg	2441
Db	75105	GGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	75164
QY	2442	-----ArgThrProLeuThrAsnArgValTyrGluAspArgProSerSerAla	2457
Db	75165	TTGGCTAAACCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	75224
QY	2458	GlySerThrProPheProTyrtAsnProLeuIleMetArgLeuGlnAlaGlyValMetAla	2477
Db	75225	GGGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	75282
QY	2478	Ser-----ProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHis	2496
Db	75283	AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	75342
QY	2497	AlaTrpAspGluGluPro	2502
Db	75343	AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	75360
RESULT 61			
AF124821			

Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.edu
----- Summary Statistics -----
Center Project name: H_NH1399P15

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Taten, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

Uncertain bases from base 63497 to base 63498.

Uncertain bases from base 186010 to base 186011.

The sequence from base 942 to base 1018 was derived from one plasmid subclone.

The sequence from base 134927 to base 135119 was derived from one plasmid subclone.

There is an unresolved tandem repeat from base 1359 to 12000.

This sequence is the entire insert of the clone.

FEATURES
source
1..192181
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/map="2"
/chromosome="2"
/clone="RP11-1399P15"
/clone_lib="RPCI-11"
2..100
/rpt_family="Alu"
repeat_region
101..160
/rpt_family="Alu"
repeat_region
161..376
/rpt_family="Alu"
repeat_region
377..436
/rpt_family="Alu"
repeat_region
437..647
/rpt_family="Alu"
repeat_region
648..702
/rpt_family="Alu"
repeat_region
703..919
/rpt_family="Alu"
repeat_region
920..970
/rpt_family="Alu"

2413 LysAlaLysSerProAlaProGly-----LeuAlaSerGlyAspArgProSerVal 2430
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1279 AGTCTAAATCTCTTCTGCGCAGAACTACTTAGGAAGTAAAGGCTTCTTCTGTC 1338
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2431 SerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrp 2450
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1339 TCCTCTGTACATTACAGAGGCTGATTACACAGCGACACACCA-----GGATGGGCTTGG 1392
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2451 GluAspArgProSerSerAlaGlySerThrProPheProTyAsnProLeuIleMetArg 2470
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1393 GAAGACAGCGCGCTTTCACAGAGCTCCACTCAGTCCCTTACACCTCTGACCATCCGG 1452
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2471 LeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGly---- 2489
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1453 ATG-----CTGACGATACCCCGCAACCCGATCGCTGTGCCCCGTGACGCCATC 1503
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2490 ProLeuAlaGlyProHis-----HisAlaTrpAspGluGluProLysProLeuLeu 2506
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1504 ACCAGCGAGCTCCGATCAACAGCGCGCATCTGGAGCGGAGCGCCCGCTCCCTC 1563
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2507 CysSerGlnTrpGluThrLeuSerAspSerGlu 2517
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1564 TCAGCACAGTATGAGACACTGTCTGACAGTGAC 1596
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 62
AC133644 192181 bp DNA linear PRI 23-NOV-2002
LOCUS
DEFINITION Homo sapiens BAC clone RP11-1399P15 from 2, complete sequence.
AC133644
AC133644 GI:25141136
VERSION
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 192181)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
9847074
2 (bases 1 to 192181)
Kim, K., Bielicki, L. and Creason, K.
The sequence of Homo sapiens BAC clone RP11-1399P15
Unpublished (2001)
REFERENCE
3 (bases 1 to 192181)
Waterston, R.H.
Direct Submission
Submitted (16-SEP-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 192181)
Waterston, R.H.
Direct Submission
Submitted (13-NOV-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 192181)
Waterston, R.H.
Direct Submission
Submitted (23-NOV-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 192181)
Waterston, R.H.
Direct Submission
Submitted (23-NOV-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 21, 2002 this sequence version replaced gi:24943062.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC

```
/rpt_family="Alu"
971. .1199
/rpt_family="Alu"
1201. .1378
/rpt_family="Alu"
1379. .1422
/rpt_family=" (CAGC) n"
1423. .1799
/rpt_family=" (CAGC) n"
1809. .1985
/rpt_family=" (GCTG) n"
2029. .2332
/rpt_family=" (GCTG) n"
2347. .2541
/rpt_family=" (GCTG) n"
2604. .2841
/rpt_family=" (CAGC) n"
2871. .3445
/rpt_family=" (CAGC) n"
3494. .3876
/rpt_family=" (CAGC) n"
3914. .4381
/rpt_family=" (CAGC) n"
4399. .4694
/rpt_family=" (CAGC) n"
4743. .4940
/rpt_family=" (CAGC) n"
4978. .5165
/rpt_family=" (CAGC) n"
5232. .6523
/rpt_family=" (CAGC) n"
6552. .6742
/rpt_family=" (CAGC) n"
6755. .7032
/rpt_family=" (CAGC) n"
7076. .7260
/rpt_family=" (CAGC) n"
7280. .7325
/rpt_family=" (CAGC) n"
7337. .7710
/rpt_family=" (CAGC) n"
7724. .8463
/rpt_family=" (CAGC) n"
8475. .9054
/rpt_family=" (CAGC) n"
9072. .9378
/rpt_family=" (CAGC) n"
9390. .9759
/rpt_family=" (CAGC) n"
9811. .10699
/rpt_family=" (CAGC) n"
10726. .11018
/rpt_family=" (CAGC) n"
11031. .12452
/rpt_family=" (CAGC) n"
12464. .12738
/rpt_family=" (CAGC) n"
14967. .15276
/rpt_family="Alu"
15381. .15437
/rpt_family="L2"
15571. .16038
/rpt_family="MaLR"
16203. .16311
/rpt_family="MaLR"
16451. .16573
/rpt_family="MaLR"
17244. .17443
/rpt_family="MaLR"
17458. .17527
/rpt_family=" (ATG) n"
17962. .18098
/rpt_family="Alu"

repeat_region 18335. .18485
/rpt_family="Alu"
repeat_region 18820. .18918
/rpt_family="L2"
repeat_region 19570. .19740
/rpt_family="L1"
repeat_region 19741. .20036
/rpt_family="Alu"

Alignment Scores:
Pred. No.: 2.22e-05 Length: 192181
Score: 709.50 Matches: 596
Percent Similarity: 33.88% Conservative: 351
Best Local Similarity: 21.32% Mismatches: 1186
Query Match: 5.37% Indels: 674
DB: 9 Gaps: 115

US-09-522-753-5 (1-2517) x AC133644 (1-192181)
Qy 15 ThrGluProArgTyrProHisSerLeuSerTyrProValGlnIleAlaArgThrHis 34
Db AGCCAGCCAGATGCCCAAGCCACCCAGTCNAAG---CCAGCTAAGCCAGCCAGCCAGCCAGCA 3510
Qy 35 ThrAspValGlyLeuLeuGluTyrGlnHisSerArgAspTyrAlaSerHisLeuSer 54
Db AGCCAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 3570
Qy 55 ProGlySerIleIleGlnProGlnArgArgProSerLeuLeuSerGluPheGlnPro 74
Db CAGGGAGCC-----AAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 3571
Qy 75 GlyAsnGluArgSerGlnGluLeuHis-----LeuArgPro 86
Db GGGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 3625
Qy 87 GluSerHisSerTyrLeuProGluLeuGlyLeuSerGluMetGluPheIleGluSerLys 106
Db AGCCAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 3685
Qy 107 ArgProArgLeuGluLeuLeuPro-----AspProLeuLeuArgProSer---- 121
Db AACCCAGCCAG-CAAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 3745
Qy 122 -----ProLeuLeuAlaThrGlyProAlaGlySerGluAspLeuThrLysAsp 138
Db CCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 3804
Qy 139 ArgSerLeuThrGlyLysLeuGluProValSerProProSerProHisThrAspPro 158
Db -----AAACCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 3840
Qy 159 GluLeuGluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArg 178
Db AGC-----CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 3879
Qy 179 ValAspArgGluIleThrMetValGluGlnGlnIle----- 190
Db -----ACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 3909
Qy 191 -----Ser 191
Db GCCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 3954
Qy 192 LysLeuLysLysLysGlnGlnGlnLeuGluGluGluAlaLysProGluProGlu 211
Db GCCTGCCAGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 4014
Qy 212 LysProValSerProProProIleGluSerLysHisArgSerLeuValGlnIleIleTyr 231
Db AAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 4074
Qy 232 AspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGlyLeuGlyProGln 251
Db -----
```


Qy 937 euLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAlaSerP 957
Db TAGCCAAAGCCA-----GCCAAGCCACCGCGCCAGCCAAAGCCAGCCAGCCAAACC 5987
Qy 957 roGlnLysProLeuAsp-----LeuLysGlnLeuLysGlnArgAlaAla- 971
Db CAGCCCAAGCCAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6047
Qy 972 -----AlaIleProProIleGlnValThrLysValHisG 983
Db AGTCAAGCCGCGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6107
Qy 983 luPro-ProArgGluAspAlaPro-----ThrLysProAla 995
Db AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6167
Qy 996 ProProAlaProProProGlnAsnLeuGlnProGluSerAspAla-----ProGln 1013
Db AAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6227
Qy 1014 GlnProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGlu 1033
Db CAGCCA---AGCCACCCAGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6281
Qy 1034 AlaPheAlaAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeu 1053
Db -----GCCAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6335
Qy 1054 ProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspProSer 1073
Db AGCCAAACAGCCAAAGCCA-----GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6383
Qy 1074 AlaPheSerTyraProProGlyHisProLeuProLeuGlyLeuHisAspThr----- 1091
Db GCCAGCCACTCAAGCCATCCAGACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6443
Qy 1092 AlaArgProValLeuProArgProProThrIle-----SerAsnProProProLeuIle 1109
Db GCAAAGCCAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6503
Qy 1110 SerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyWet 1129
Db AGCCAAAGCCGCGCGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6563
Qy 1130 SerValGlnLeuHisValProTyraSerGluHisAlaLysAlaProValGlyProValThr 1149
Db CAC-----CCGCGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6605
Qy 1150 MetGlyLeuProLeuProMetAsp-----ProLysLysLeuAlaProPhe 1164
Db ACAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6665
Qy 1165 SerGlyValLysGlnGluLeuSerProArgGlyGlnAlaGlyProProGluSerLeu 1184
Db GCGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6725
Qy 1185 GlyValProThrAla-----GlnGluAlaSerValLeuArgGlyThrAlaLeuGly 1201
Db AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6785
Qy 1202 SerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspSer 1221
Db GCGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6845
Qy 1222 AlaIleThrTyraArgGlySerIleThrHisGlyThrProAlaAspValLeuTyLysGly 1241
Db GCC----- 6848
Qy 1242 ThrIleThrArgIleIleGlyGluAspSerProSerArg---LeuAspArgGlyArgGlu 1260
Db -----AACCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6884
Qy 1261 AspSerLeuProLysGlyHisValIleTyraGluGlyLysLysGlyHisValLeuSerTyra 1280

Db 6885 GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6911
Qy 1281 GluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyPro 1300
Db -----CAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6944
Qy 1301 ProHisGluThrAlaAlaProLysArgThrTyraAspMetMetGluGlyArgValGlyArg 1320
Db -----AACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7001
Qy 1321 AlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHis 1340
Db GGCATCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7049
Qy 1341 SerProHisHisLeuLysGlu----- 1347
Db TCACCCCACTCAAACTAGCCCAAGTCACCCGCGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 7109
Qy 1348 -----GlnHisHisIleArgGlySerIleThrGlnGlyIleProArgSer 1362
Db CCCAGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7168
Qy 1363 TyrValGluAlaGlnGluAspTyraLeuArgArgGluAlaLysLeuLysArgGluGly 1382
Db AGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7219
Qy 1383 ThrProProProProProProProProProProProProProProProProProProProProPro 1401
Db -----CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7273
Qy 1402 LeuGlyProLeuLysLeu-----LysProAlaHisGluGlyLeuValAlaThr---Val 1418
Db ATCAAAACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7333
Qy 1419 LysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGlu 1438
Db AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7393
Qy 1439 LeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLys 1458
Db AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7453
Qy 1459 TyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuIleGly 1478
Db CCAGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7501
Qy 1479 SerProGlyArg-----ThrPheProProValHisPro 1489
Db CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7561
Qy 1490 LeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyraGluGluSerLeuLys 1509
Db GG-CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7620
Qy 1510 SerArgProGlyThrAlaSerSerSerGlySerIleAlaArgGlyAlaProValIle 1529
Db AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7671
Qy 1530 ValProGluLeuGlyLysProArgGln-----SerProLeuThrTyraGluAspHisGly 1547
Db -----GGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7716
Qy 1548 AlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrPro 1567
Db -----CCAAGA---GAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7751
Qy 1568 ArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSer 1587
Db ----- 7752
Qy 7752 CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7799
Qy 1588 ThrProArgGluIleAlaLysSerProHisSerThrValProGluHisHisProHisPro 1607


```

Qy 2208 -----GluproAsnLysT 2212
Db 9831 CAAGCCGCCACCCATCAAGCCAGCAAGCCAGCAAAACCCAGCAGCCAGCCAGCCAGCA 9890
Qy 2212 hrSerValLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 2232
Db 9891 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 9941
Qy 2232 hrGluProGlyHisSerArgSerAlaValTyProLeuLeuTyArgAspGlyGluGlnT 2252
Db 9942 CCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 9989
Qy 2252 hrGluProSerArgMetGlySer-----LysSerProG 2263
Db 9990 GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAA 10049
Qy 2263 LysThrSerGlnPro-----ProAlaPhePheSerLysLeuThrGluS 2278
Db 10050 ACCAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 10109
Qy 2278 erAsnSerAlaMetValLysSerLysLysGlnGluLeuAsnLysLeuAsnThrHisA 2298
Db 10110 AGTCAGC-----CAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 10157
Qy 2298 snArgAsnGluProGluTyAsnLysSerGlnProGlyThrGluLeuPheAsnMetPro- 2317
Db 10158 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAA 10211
Qy 2318 AlaIleThrGlyThrGlyLeuMetThrTyArgSerGlnAlaValGlnGluHisAlaSer 2337
Db 10212 GCAAGCCATGGAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAA 10271
Qy 2338 ThrAsnMetGlyLeuGluAlaIleAlaArgLysAlaLeuMetGlyLysTyArgGlnTirp 2357
Db 10272 ACCAGCAAGCCAGCCAA-----CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 10319
Qy 2358 GluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeu 2377
Db 10320 AGCCAGTCAGCCAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 10373
Qy 2378 ProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerPro 2397
Db 10374 CCAGCCAGCCAGCCAA-----AGCCGCGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGAT 10418
Qy 2398 GlyGlyGlyGlyLys---AlaLysValSerGlyArgProSerArgLysAlaLysSer 2416
Db 10419 GGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 10478
Qy 2417 ProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGlu 2436
Db 10479 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 10538
Qy 2437 GlyAspCysAsnArgArgThr---ProLeuThrAsnArgValTrpGluAspArgProSer 2455
Db 10539 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 10598
Qy 2456 SerAlaGlySerThrProPheProTyAsnProLeuIleMetArgLeuGlnAlaGlyVal 2475
Db 10599 CAGTCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 10657
Qy 2476 MetAlaSerProProProGlyLeuProAla---GlySerGlyProLeuAlaGlyPro 2494
Db 10658 GTCACTCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 10716
Qy 2495 HisHieAlaTrpAspGluGluPro---LysPro 2504
Db 10717 AGA-----TGGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 10743

RESULT 63
BX248336/c
LOCUS
DEFINITION Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment 3/14.

```

```

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL

REMARK
REFERENCE
AUTHORS
TITLE
JOURNAL

FEATURES
source

gene
CDS

```

```

BX248336 BX248333
BX248336.1 GI:31617341
complete genome.
Mycobacterium bovis subsp. bovis AF2122/97
Mycobacterium bovis subsp. bovis AF2122/97
Mycobacterium bovis subsp. bovis AF2122/97
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
1
Garnier, T., Eiglmeier, K., Camus, J.-C., Medina, N., Mansoor, H.,
Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsemp, C., Simon, S.,
Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L.,
Wheeler, P.R., Parkhill, J., Barrell, B.G., Cole, S.T., Gordon, S.V. and
Hewinson, G.
The complete genome sequence of Mycobacterium bovis
Online Publication
PNAS 10.1073/pnas.1130426100 ( Microbiology )
2 (bases 1 to 320050)
Garnier, T.
Direct Submission
Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire
Bacterienne Institut Pasteur 28 rue du Dr Roux 75724 PARIS cedex
15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the
Mycobacterium Bovis sequencing teams, TB Research Group, Veterinary
Laboratories Agency, Weybridge, Woodham Lane, New Haw, Addlestone,
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK. PT4 Annotation, Genopole, Institut
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28
rue du Docteur Roux, 75724 Paris Cedex 15, France
Location/Qualifiers
1. 320050
/organism="Mycobacterium bovis subsp. bovis AF2122/97"
/mol_type="genomic DNA"
/strain="AF2122/97"
/db_xref="taxon:233413"
complement(93..1484)
/locus_tag="Mb0588c"
complement(93..1484)
/locus_tag="Mb0588c"
/notes="Mb0588c, -, len: 463 aa. Equivalent to RV0573C,
len: 463 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 463 aa overlap). Conserved
hypothetical protein, similar to other conserved
hypothetical proteins and some nicotinate
phosphoribosyltransferases e.g. NP_213718.1|NC_000918
hypothetical protein from Aquifex aeolicus (425 aa);
AL109962|T36953|SCU1.20 conserved hypothetical protein
from Streptomyces coelicolor (438 aa), FASTA scores: opt:
1089, E(): 0, (49.4% identity in 385 aa overlap);
P_391053.1|299120|BSUB0017_57|NC_000964 protein similar to
nicotinate phosphoribosyltransferase from Bacillus
subtilis (490 aa), FASTA scores: opt: 955, E(): 0, (43.5%
identity in 356 aa overlap); etc. Also similar to
Q10641|Y03F.MYCU|MTCY130.15c|RV1330c CONSERVED
HYPOTHETICAL PROTEIN from Mycobacterium tuberculosis (509
aa), FASTA scores: opt: 761, E(): 0, (38.4% identity in
437 aa overlap)."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD93450.1"
/db_xref="GI:31617342"
/translation="MAIROHVGALFTDLVETMAQYWAERMSGTAVFETPFKRLPGG
RSYIMAGLADVVEFLAEFRFQDLRLYLGLQFSEDFLRLWLAGRVFTGDVWAPG
TVTFNPEPAVLIIAEQLVETFINQIHLQSVLAKAAVVAARVAGRPVDFGAR
RHGTDACKVARTSYLAGAGTSNLLAARQYGIPTGTWHSFVQAFDSVAFEARL
ARLYPATMLLVDTYTLRGVDHVLKLNRFVRAVRLDGLDGLDLSKATRLD
TAGLEQVEIPASSGLDENRIALLAARCPIDGFGVGTQLVVAODPALDMAYKLAVD
GSRRTFSSGKVIYPCRQVFRKLEHGVCGDPLGSHGENLPGDPLLMVYTWGRRLR
QHAPTLDGARDWARQQIDALPPELSLEDTGTSYPVAVSDRIVGELARLRHADTAEH
PGSNVVGAKARP"
complement(1494..2636)

```



```
Qy 728 yAsnGluValProArgGlyGluCysSerGlyProAla----- 740
Dy 175585 CGGTCAAGCTGGAACCGCGCGTGGCGCGCGCGCGCGAGGAGCAGC 175526
Qy 741 -----ThrValAsnAsnSerSerAspThrGluSer-IleP 752
Dy 175525 CGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCTATGGCTGC 175466
Qy 752 roSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlat 772
Dy 175465 CGGCGCGCGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 175406
Qy 772 hrLeuGlyAlaAspGlyProProProGlyProProThrProProArgThrSer---- 790
Dy 175405 CGCGCGCGAGGGTCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 175346
Qy 791 -----ArgAlaProIleGluProThrProAlaSer-----GluAlaThrGlyA 805
Dy 175345 CGTTCCACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 175286
Qy 805 laProThrPro----- 808
Dy 175285 CGCGCGTCCCGGTGGCGAAGCGCGCGCGGTTCACCGCGCGCGCGCGCGCGCGCG 175226
Qy 809 -----ProProAlaProProSerProSerAlaPro----- 818
Dy 175225 CGCGCGCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 175166
Qy 819 -----ProProValValProLysGluL 827
Dy 175165 CGCGCGCGCGCGCATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 175106
Qy 827 ysGluGluGluThrAlaAla----- 833
Dy 175105 AGCGGAGACGACGCGCGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 175046
Qy 834 -----AlaProValGluGluGlyGluGluGlnLysPro-----P 846
Dy 175045 GCGCGCGCGTCCCGCGGTGCC-GCGGTGCCGAACAAGATCCCGCGCGCGCGCGCG 174987
Qy 846 roAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluC 866
Dy 174986 CGCGCGCGCGCG-----GCGGTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 174933
Qy 866 ysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaL 886
Dy 174932 TTCCGGAACAA-----CCAGCGCGCGGTCCCGCGCGCGCGCGCGCGCGCG 174885
Qy 886 hrAlaGluGlyAlaLeu-----LysAlaGluLysLysGluGly----- 898
Dy 174884 GTCCCGTTGGCGCGTTCGCGATCAGCGGACGCGCGGTGAGCGTCTGGAAGGGTCTTC 174825
Qy 899 -----GlySerGlyArgA 903
Dy 174824 ACCATATTGATCATCGTCTGCGAGGTGTGCGAGGGCGAGGTGTGGGGGGCGGTA 174765
Qy 903 laThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerAlaThrCysSerA 923
Dy 174764 AAGCGTCCAGGGCCACACACACCGCGCGCGGTACCGCGCGATGCC----- 174721
Qy 923 laAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeu-LeuSerProArgPro 942
Dy 174720 -----GGCTTCCCGCGAGGTCCGCGCGGTCCCGCGCGCGCGCGCGCG 174684
Qy 943 SerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAsp 962
Dy 174683 CCGTTG-----CCGATCAACAGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCG 174630
Qy 963 LeuLysGlnLeuLysGlnArgAlaAlaAlaProProIleGlnValThrLysValHis 982
Dy 174629 CGCGCGGTGTCGCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCG 174573
```

Db	173700	GTGGTCCGGTCCCGGGCGCGCTTGGCGCGTTCGCCGATCAGCGGGCGCGCGGTGGC	173641	Db	172884	AGCAACCCGCGCGCGCGCGCTTCCGGGGTTCGGCGG	172884
Qy	1339	ArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGly	1358	Qy	1656	AlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAsp	1675
Db	173640	CGCACGAATTGGCGTGTGATCGGGCCAGCAGCGCGACACGGC	173593	Db	172842	GGCGGGTGGCGCGCTTCCCGCGTTCGGCGGTGCGATCAGCCCGCGCGTCCGCGG	172783
Qy	1359	IleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeu	1378	Qy	1676	ThrAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMet	1695
Db	173592	GG-CCTCGG	173564	Db	172782	ACGCCACCGCGCGCG	172768
Qy	1379	LysArgGluGly	1389	Qy	1696	HisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerPro	1715
Db	173563	CGCGCTCAGGGCTGCACGAATGGCGGTAAACCGCGCGCTGTGCGTCCGCGCT	173504	Db	172767	-----CCGGTGTGCGCG	172756
Qy	1390	SerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGly	1405	Qy	1716	ArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSer	1735
Db	173503	GGTAGG	173468	Db	172755	ACGCTTAAGGTGCGCGCTTCCGGCGCGCGCGCG	172717
Qy	1406	LysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIle	1425	Qy	1736	GlnValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMet	1755
Db	173467	TCGCTGTGCACACCTCATCGCACCGCGGTCAACAATCCCGTGTGACGGCGCGG	173408	Db	172716	---CTACCCCAATACTGACCGCGGTGCGCGTTCACCGCGCGCGCGTTCAGGTTA	172660
Qy	1426	HisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgPro	1445	Qy	1756	AspArgLeuAlaTyrIleuProThrAlaProGlnProPheSer-SerArgHisSerSerSe	1775
Db	173407	CCGATTAGCGAGCTAATCGCGACCCGATCCCGGCAAAATCGCTGCGCGCGCACCA	173348	Db	172659	CGCGCATTCGCGCATGCCCCCGCGCGCGCGAAGCCAGCCAGCCGCGCATTCGCGCGG	172600
Qy	1446	LeuLysGluGly---SerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSer	1464	Qy	1775	rProLeuSerPro-----GlyGlyProThrHisLeuThrLysProThrTh	1790
Db	173347	CC-AGCTCCGGCGAACCATCACC	173325	Db	172599	GCCCCACCGACCCACCACCGCGGTGCGGAACCCCGCGAACCCCGCGCGCGCGTGGAC	172540
Qy	1465	ThrThrGlySerLysLysHisAspValArgSerIleuLeuGlySerProGlyArgThrPhe	1484	Qy	1790	rThrSerSerSerGluArgGluArgAspArgAspArgGluArgAsp-----ArgAs	1807
Db	173324	-----CATGATCGCATGCTC	173307	Db	172539	AACAGCAGCCCGCGGTGCGCACCGCGCGCTCCACCGCGCGTCCACCGAGAAACCAATCCGCGG	172480
Qy	1485	ProProValHisProLeu-----AspValMetAlaAspAlaArgAlaLeuGluArgAla	1502	Qy	1807	pArgGluArgGluLysSerIleLeuThrSerThrThrThrValGlu-----HisAlaPr	1825
Db	173306	CCAAACCTCAGCGCTGTCGCGCATACACGCGCACACCCGCTAGC---GGTAGGAAT	173250	Db	172479	GGCGCGCGCGCGCGCGTGCAGAACAGCAGCCCTCGCTACCGCGCGCGCGCGCGTCC	172420
Qy	1503	CysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSer	1517	Qy	1825	oileTyrArgProGlyThrGluGln-SerSerGlySerSerGlySerSer-----GlyG	1843
Db	173249	TAATTGCAGAAATCAACGATATCGGATCCGGCACCGCATGCGCTTCCACACGCTTC	173190	Db	172419	CMACCGGTTCAGGTTTTCGCGTTCACCGCGCTCCCGCGCGCGCGCGCGCGCGG	172360
Qy	1518	SerGlyGlySer-----IleAlaArgGlyAlaProValIleValProGluLeuGlyLys	1535	Qy	1843	lyGlyGlyGlySerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProI	1863
Db	173189	GCAGCGCGCAGCTGCTGACGGCTACGGGTGCGG-CGGCGGTTCGCGAACACCGCGCG	173131	Db	172359	AGCATTCGCGCATTCGCGCGCGCGCGCAACCCCGCGCTCGCCATTAGTTGTAGCCCG	172300
Qy	1536	ProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuPro	1555	Qy	1863	leSerProArgThrGlnAspAlaLeuGlnArgProSerValLeuHisAsnThrGlyM	1883
Db	173130	CGGTCCCGGCA-----CAGGGCGCGCGCGAACCAGCGAC-----	173095	Db	172299	CCGGTCCCGCGTGACCTCCCGCGGACCGATCAGCCCGCGC---CAGCAGTCCCGCGGTC	172243
Qy	1556	ArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSer	1575	Qy	1883	etLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSert	1903
Db	173094	-----CGCGCGGATAGCGTCTGCACGACGACCGCGA-----CGGGCC	173056	Db	172242	CCGCGCGC-----CCACCGCGCGCGCGCTC-----GCCA	172213
Qy	1576	SerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSer	1595	Qy	1903	hrSerSer-ProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGly	1922
Db	173055	TCTGAAAGCGCTCGAGCGCGG-----GGAICTTCGCCGAGATACCTGCGCGCGCGG	173002	Db	172212	CCAAAGCGCGGTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	172159
Qy	1596	Pro-----	1596	Qy	1923	GlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAla	1942
Db	173001	CCCCATTTTGGCGCGCTCCGCGCGCGATCTGCATCACCAGCGAGTGTGACGCGC	172942	Db	172158	-----CCGACCCCGCTCCCGCGCGCGCGCGAGT-----	172126
Qy	1597	---HisSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeu	1615	Qy	1943	ProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLys	1962
Db	172941	GGAAGCCATGCGTCCCGCGCTACCGCAACCGTTGAGCCCG	172897	Db	172125	---ACGGTGTGCTCCGCGATCCCGCGCGC-----	172096
Qy	1616	ArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThr	1635	Qy	1963	ProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArg	1982
Db	172896	-----TTTGGCCGACG	172885	Db	172095	CCGCGCGTCG-----CCGAGCAGCGACGCGC-----CCGGCC	172065
Qy	1636	SerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeu	1655	Qy	1983	ProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeu	2002
				Db	172065	CCGCGCGCGCGCGCGGTACCC---CACAGCGCG---GCAGCCCGCGCGCGCGCGCGG	172012


```
REFERENCE
AUTHORS
3 (bases 1 to 200389)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Maclean,C., Macdonald,P., Major,J.,
Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Mianga,V., Murphy,T., Naylor,J., O'Donnell,P., O'Neil,D., Oliver,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (11-DEC-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 200389)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Maclean,C., Macdonald,P., Major,J.,
Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Mianga,V., Murphy,T., Naylor,J., O'Donnell,P., O'Neil,D., Oliver,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-JAN-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 16, 2003 this sequence version replaced gi:26449133.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIGR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L28245
Center clone name: 199_M_12
-----
Location/Qualifiers
1..200389
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone_lib="RPIC-13 Human Female BAC"
1881..2390
/rpt_family="LOR1a"
complement(2549..2616)
/rpt_family="MIR"
complement(2782..2902)
/rpt_family="L2"
complement(3047..3272)
repeat_region
repeat_region
repeat_region
repeat_region

/rpt_family="MIR"
complement(3553..3716)
/rpt_family="MIR"
complement(3849..3912)
/rpt_family="MLT1D"
3913..4179
/rpt_family="AluSc"
complement(4180..4615)
/rpt_family="MLT1D"
complement(4630..4708)
/rpt_family="L2"
complement(4725..4802)
/rpt_family="L2"
complement(5008..6173)
/rpt_family="L1PA3"
7787..7914
/rpt_family="L2"
8285..8348
/rpt_family="AluY"
8569..8611
/rpt_family="CT-rich"
8612..8642
/rpt_family="(CA)n"
8657..8791
/rpt_family="(CA)n"
8827..8878
/rpt_family="(CA)n"
9186..9484
/rpt_family="AluSx"
complement(9842..9983)
/rpt_family="L2"
10727..10981
/rpt_family="MIR"
11251..11705
/rpt_family="MLT2B1"
complement(11833..12418)
/rpt_family="MER82"
complement(12948..13100)
/rpt_family="L1M4"
complement(13219..14843)
/rpt_family="L1PA8"
14847..15616
/rpt_family="L1P"
complement(15907..16160)
/rpt_family="HERV16"
complement(16198..16311)
/rpt_family="HERV16"
complement(16319..16620)
/rpt_family="AluY"
complement(17103..17204)
/rpt_family="MER20"
17668..17813
/rpt_family="MIR"
17838..17880
/rpt_family="(TA)n"
18877..19017
/rpt_family="MIR3"
19402..19424
/rpt_family="AT-rich"
19638..19658
/rpt_family="AT-rich"
20031..20100
/rpt_family="L2"
20638..20704
/rpt_family="AT-rich"
complement(21399..21823)
/rpt_family="L1MB4"
complement(21829..22010)
/rpt_family="L1PA10"
22011..22212
/rpt_family="L2"
complement(23550..25788)
/rpt_family="L1MD2"
```


QY	1876	erValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValAlaGluProSerLysProT	1896
DB	47527	CC-----AAGCCAGCCAAGCAAGCACCACCGCAAGCACGCCA--	47490
QY	1896	hrValLeuArgSerThrSerSerProValArgProAlaAlaThrPheProProA	1916
DB	47489	-----ACCAGAG-AGCCAGGCAATTCAGCCCTAGACAGCCAAAGCAAGCCAG	47445
QY	1916	laThrHisCysProLeuGlyGlyThrLeuAspGlyValTyProThrLeuMetGluProv	1936
DB	47444	CCAGTCATGCC-----	47434
QY	1936	alLeuLeuProLysGluAlaProArgValAlaArgProGluUArgProHrgAlaAspThrG	1956
DB	47433	-----AGCCAGCAGCCCAAGTCAGCCAGCCCAAGCA--	47398
QY	1956	lyHisAlaPheLeuAlaLysPro-ProAlaArgSerGlyLeuGluProAlaSerSerPro	1975
DB	47397	-----GCCAAGCCAGCCAGCCAGCCAGCCAGCCAAATCCAGCCAGCCACCCA	47352
QY	1976	SerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAla	1995
DB	47351	AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	47295
QY	1996	ArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaProPro	2015
DB	47294	ACCCAGCCAGCCAGCC-----CAGCCGAAGCCAGCCAAACAGCCAGCCAGCCAA	47247
QY	2016	AlaSer--AlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGln	2034
DB	47246	GCCAGCCAAAGTCACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	47190
QY	2035	GluLeuGluLeuArgSerLeuGlyTyThrHisGlySerSerTySerProGluGlyValGlu	2054
DB	47189	CCAGCCAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	47130
QY	2055	ProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeu	2074
DB	47129	CCACTGAAGCCAGCCAGCCAGCCAGCCAGCC-----	47103
QY	2075	GluGluLeuAspLysSerHisLeuGluGlyLeuArgProLysGlnProGly---Pro	2093
DB	47102	-----CAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	47079
QY	2094	ValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProGluSerGln	2113
DB	47078	ACCAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	47025
QY	2114	ProSerSerSerPro-LeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValva	2133
DB	47024	CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAA	46973
QY	2133	lThrLeuAlaGlnHisLysSerGluValIleThrGlnAspTyThrArgHisHisPro--	2152
DB	46972	-----GCCAGCCAAAGCCAAACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	46929
QY	2153	-----GlnGlnLeuSerAlaProLeuProAlaProLeuTySerPh	2166
DB	46928	CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	46869
QY	2166	eProGlyAlaSerCysProValLeuAspLeuArgProProSerAspLeuTyLeuPr	2186
DB	46868	CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	46821
QY	2186	oProProAspHisGlyAlaProAlaArgLysSerProHisSerGluGlyLysArgSe	2206
DB	46820	AAG-CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	46762
QY	2206	rProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValse	2226
DB	46761	CCAGGCACCCAAACCAACCAAGC-----CAGCCAGCCAGCCAGCCAGCCAG	46729
QY	2226	r-----ProProGluGlyMetThrGluProGlyHisSerArgSe	2239

[illegible]

RESULT 65	VERSION
SCO939111	KEYWORDS
LOCUS	SOURCE
DEFINITION	
ACCESSION	

ORGANISM Streptomyces coelicolor A3(2)
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE
 AUTHORS Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C. H., Kieser, T., Larke, L., Murphree, L., Oliver, K., O'Neill, S., Rabinowitz, E., Rajandream, M. A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorek, A., Woodward, J., Barrell, B. G., Parkhill, J., and Hopwood, D. A.

TITLE Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)

JOURNAL Nature 417 (6885), 141-147 (2002)

MEDLINE 21996410

PUBMED 12000953

REFERENCE 2 (bases 1 to 321250)

AUTHORS Bentley, S. D.

TITLE Direct Submission

JOURNAL Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sds@sanger.ac.uk

COMMENT On or before Oct 29, 2002 this sequence version replaced gi:20520791, gi:20520768, gi:20520769, gi:20520793, gi:20520794, gi:20520822, gi:20520770, gi:20520690, gi:20520671, gi:20520801, gi:20520696, gi:20520807, gi:20520866, gi:20520790.

FEATURES
 source
 1..321250
 /organism="Streptomyces coelicolor A3(2)"
 /mol_type="genomic DNA"
 /strain="A3(2)"
 /db_xref="taxon:100226"
 complement(200..2215)
 /gene="SC01964"
 /note="synonym: SCC54.24c"
 complement(200..2215)
 /gene="SC01964"
 /note="SCC54.24c, possible export associated protein, len: 671aa; prediction based on the predicted function of surrounding genes and the short region (67-162aa) of similarity to TR:Q52357 (EMBL:U38824) terD, associated with phage inhibition, colicin resistance and tellurite resistance from Serratia marcescens IncHI2 plasmid R478
 blast scores: E=2.1e-12, Identities = 40/98 (40%)
 Positives = 55/98 (56%) which may represent an active site."
 /codon_start=1
 /transl_table=1
 /product="putative export associated protein"
 /protein_id="CAB38154.1"
 /db_xref="GI:4468700"
 /db_xref="GOA:Q9Z502"
 /db_xref="SPTREMBL:Q9Z502"
 /translation="MTAELTRGQNHPLPRARLEIRVSAGTPVVGATLPGDNGTTHGV
 ERVHPGAPTLPGLEVSROAAHRLAVALDVPDAVHRVSVLLALPAGSGPARFCA
 VAAVFPATLDGDGVASYITGLAESAVVALELVRQAGKVRVAGGVAGYGLAEI
 LAQKLSQAHLAATHEAVASGLARISIPAPPAAPNRRPDHGMFGAGGPGVPFAS
 FYDPGPPAPAPGPPGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG
 FADSRMEKELDPSDRISIGGGDAAREARHGLVQSAREVLDLDRDVAQLVAEA
 EVVPEALPAAPAWNDVNHAYRVPMEIPWALRLGLDHLPEADRIIPMLIRIPLBERG
 LWIDSGRSASLGSFADSHMRRLGLETAVSAAARLLAVYPAGEFTVHVTDPAQSGAQ
 ALAPLAGSVLAPPAQAGADVLARVQRVDVLQVMAALRGAPADLPGLDTSQQL
 LIYVDFPHGDDRAVNGYAGADVAVGVHLMWADREESAGFGPLLDPLWRSLRL
 TPVADHDLDPWGHAWTYEPFLVPPGSGVQLVQVLAQVAARSRWDR"
 complement(2340..2318)
 /gene="SC01965"
 /note="synonym: SCC54.25c"
 complement(2340..2318)
 /gene="SC01965"
 /note="SCC54.25c, possible export associated protein, len: 671aa; prediction based on the predicted function of surrounding genes and the short region (67-162aa) of similarity to TR:Q52357 (EMBL:U38824) terD, associated with phage inhibition, colicin resistance and tellurite resistance from Serratia marcescens IncHI2 plasmid R478
 blast scores: E=2.1e-12, Identities = 40/98 (40%)
 Positives = 55/98 (56%) which may represent an active site."
 /codon_start=1
 /transl_table=1
 /product="putative export associated protein"
 /protein_id="CAB38154.1"
 /db_xref="GI:4468700"
 /db_xref="GOA:Q9Z502"
 /db_xref="SPTREMBL:Q9Z502"
 /translation="MTAELTRGQNHPLPRARLEIRVSAGTPVVGATLPGDNGTTHGV
 ERVHPGAPTLPGLEVSROAAHRLAVALDVPDAVHRVSVLLALPAGSGPARFCA
 VAAVFPATLDGDGVASYITGLAESAVVALELVRQAGKVRVAGGVAGYGLAEI
 LAQKLSQAHLAATHEAVASGLARISIPAPPAAPNRRPDHGMFGAGGPGVPFAS
 FYDPGPPAPAPGPPGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG
 FADSRMEKELDPSDRISIGGGDAAREARHGLVQSAREVLDLDRDVAQLVAEA
 EVVPEALPAAPAWNDVNHAYRVPMEIPWALRLGLDHLPEADRIIPMLIRIPLBERG
 LWIDSGRSASLGSFADSHMRRLGLETAVSAAARLLAVYPAGEFTVHVTDPAQSGAQ
 ALAPLAGSVLAPPAQAGADVLARVQRVDVLQVMAALRGAPADLPGLDTSQQL
 LIYVDFPHGDDRAVNGYAGADVAVGVHLMWADREESAGFGPLLDPLWRSLRL
 TPVADHDLDPWGHAWTYEPFLVPPGSGVQLVQVLAQVAARSRWDR"
 complement(2340..2318)
 /gene="SC01965"
 /note="synonym: SCC54.25c"
 complement(2340..2318)
 /gene="SC01965"
 /note="SCC54.25c, possible export associated protein, len: 671aa; prediction based on the predicted function of surrounding genes and the short region (67-162aa) of similarity to TR:Q52357 (EMBL:U38824) terD, associated with phage inhibition, colicin resistance and tellurite resistance from Serratia marcescens IncHI2 plasmid R478
 blast scores: E=2.1e-12, Identities = 40/98 (40%)
 Positives = 55/98 (56%) which may represent an active site."
 /codon_start=1
 /transl_table=1
 /product="putative export associated protein"
 /protein_id="CAB38155.1"
 /db_xref="GI:4468701"
 /db_xref="GOA:Q9Z501"
 /db_xref="SPTREMBL:Q9Z501"
 /translation="MTVNTKGAISLQKSGSLTSVRMLGWAQAPRCLFGSRTR
 EIDLASAVLPADKQPDVDFVFRHLVSDSDSVRHTGNLVGVCQGGDEAILLDVLR
 VPHIDQIVFTVNSFTGTFQFQVONAFCLVDETNGQELARYTLAGGAFQAQIMAKV
 HRAGQGSMTAIGTPANGRTFQDLPAILPVL"
 complement(3061..5199)
 /gene="SC01966"
 /note="synonym: SC3C9.01c, SCC54.26c, uvrB"
 complement(3061..5199)
 /gene="SC01966"
 /note="SCC54.26c, partial CDS, probable uvrB, ABC excision nuclease subunit B, len: >113aa; similar to many eg. SW:UVRB MICLU uvrB, ABC excision nuclease subunit B from Micrococcus luteus (709 aa) fasta scores, opt: 248, z-score: 280.5, E(): 2.6e-08, (45.9% identity in 111 aa overlap).
 SC3C9.01c, uvrB, excinuclease ABC subunit B (fragment), len: >635 aa; highly similar to SW:UVRB BACSU (EMBL:AF017113) Bacillus subtilis excinuclease ABC subunit B (DinA protein) UvrB or DinA or Uvr aa; fasta scores: opt: 2591 Z-score: 2687.6 bits: 507.5 E(): 4.3e-142; 63.667% identity in 600 aa overlap. Contains Pfam matches PF00271 helicase C, Helicase conserved C-terminal domain and match to ProSite entry PS00017 ATP/GTP-binding site motif A (P-loop). Also contains a possible coiled-coil region at approx. residues 273..290"
 /codon_start=1
 /transl_table=1
 /product="ABC excision nuclease subunit B"
 /protein_id="CAD55287.1"
 /db_xref="GI:24419016"
 /db_xref="GOA:Q8CK11"
 /db_xref="SPTREMBL:Q8CK11"
 /translation="MRPVSIERTVAPFEVYVSPYQPSGQDTAIAELARRVQAGEKV
 VILGATGTGKSATTAWIEKIQRTLVMAPNKTLAAQLANEFRELLPNNAVEVFSVY
 DYYQRAYVPSQDTYIEKQSSINEEVERLHRSATNSLTRDDVIVVASVCIYGLTP
 QEYVDRMVLVCGEEHDLRLRFVDIQYTRNDMAFARGTFVRGTIIFPVEEL
 AVRIEMFGDEIALSLHVPVTEIISDQQLYVFPASHYVAGPERLRAVNDIKELA
 ERLELEKQKGLLEAQRMRRTTYIDEMIQICSGGVENYSMHFGDRCSPSPNTLL
 DYPDDFLVVIDESHVTPQIGAMYEQIDASKRTLVDFGRFLPSALNRLKWEFQE
 RIGQTVLALPGAVELSDGAVQEIIRPTGLVDPEVVKVFKPQDDLVHETRRRT
 EKDERVLTTLTKKMAEDLTDYFVGLGVYRVLHSDVDTLRRVELLELRAGEYDVILV
 GINLLREGLDLPVSLVALDADKEGFLSGTSLIQTIGRAARNVSQVHMYADKITP
 AMKAIDETNRRREKVAFNKGVDPPQPLRKKNIDVIAQIAREVDVTEQLGSGYQK
 TKEGKAQKAPVPALGGQTKGAARAKETATVTRPAAELAEQIBDLTTRMRANA
 DLQFEAAALRDEVSMMKELQWRAGLA"
 complement(3541..3801)
 /gene="SC01966"
 /note="Pfam match to entry PF00271 helicase_C, Helicase conserved C-terminal domain, score 75.00, E-value 1.5e-18"
 complement(4753..5166)
 /gene="SC01966"
 /note="Pfam match to entry PF00270 DEAD, DEAD/DEAH box helicase, score 27.00, E-value 1.2e-07"
 complement(5035..5058)
 /gene="SC01966"
 /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
 complement(5239..6117)

```
/gene="SC01967"
/notes="synonym: SC3C9.02c"
complement (5239, .6117)
/genes="SC01967"
/notes="SC3C9.02c, possible integral membrane protein, len:
292 aa; similar to TR:Q9KZ14 (EMBL:ALJ53863) Streptomyces
coelicolor putative integral membrane protein SCG8A.15c,
244 aa; fasta scores: opt: 703 Z-score: 726.3 bits: 142.1
E(): 7.6e-33; 48.430% identity in 223 aa overlap. Contains
possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAC42741.1"
/db_xref="GI:14571767"
/db_xref="SPTREMBL:Q93JJ4"
/translation="MNSTVQGCYGLVTPLVYLMACLGALGRLCTARAVLSRWR
PGWALGSAIGSIVMTMFVAMNGFTIEHTPIRYDMLTFASLAVIMVGVGIFIV
GYRGARTALFTGGTITGIGIASNHYLGAGMHLDCQLTYNTFTVAVSVVIAVATA
ALWAAGQVRGLWSVGAALIMGLAVTMGMVAEVEHLGTAEPVSGVSGPAELLAP
MLTGPLALLAGVVMFDPMLVMGRPARVPAERKPGIIPAETVEPTIRHPAHRPRR
TRPLGHRRSRTFQNR"
6399.6403
6409.7278
/genes="SC01968"
/notes="synonym: SC3C9.03"
6409.7278
/genes="SC01968"
/notes="SC3C9.03, probable secreted hydrolase, len: 289 aa;
similar to SW:GLPQ_BACSU (EMBL:Z26522) Bacillus subtilis
glycerophosphoryl diester phosphodiesterase (SC 3.1.4.46)
GLPQ, 293 aa; fasta scores: opt: 495 Z-score: 576.9 bits:
114.7 E(): 1.6e-24; 34.114% identity in 299 aa overlap.
Contains possible cleavable N-terminal region signal
peptide sequence"
/codon_start=1
/transl_table=11
/product="putative secreted hydrolase"
/protein_id="CAC42742.1"
/db_xref="GI:14571768"
/db_xref="GOA:Q93JJ3"
/db_xref="SPTREMBL:Q93JJ3"
/translation="MWRVAVVTTALLGVALTAPLSHARADAGDDPTWVAHRGAS
GYAENTLAAVDRAELGIRFVENDVQTRDGLVHLHDESLARTTVEVFDPDRSPW
KVQDFTAEIARLDAGSIFGSEYAGARVPTLSEQVHRVHNHQLLELSPGLYGI
EQOTLKVLANGLDRHRVAGRLVQSFADSIKTVHDLKPAVATGFLGTPAVSELPA
```

Alignment Scores:

Pred. No.:	4.34e-05	Length:	321250
Score:	704.00	Matches:	657
Percent Similarity:	30.98%	Conservative:	259
Best local Similarity:	22.20%	Mismatches:	1112
Query Match:	5.33%	Indels:	948
DB:	1	Gaps:	130

US-09-522-753-5 (1-2517) x SC0939111 (1-321250)

```
Qy 16 GluProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThr 35
Db 64068 CAGCGGTAGTCCGCCAGCAGCTCGTGGCCACCGCTG-----CGTGCCACAGTA 64018
Qy 36 AspValGlyLeuLeuGluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerPro 55
Db 64017 CAGGTGCTCGTGCAGAACCGCACCTCCACCGGGGCCACACCTGGCCGAGAACGTCG 63958
Qy 56 GlySerIle-----lleGlnProGlnArgArgArgProSerLeuLeuSer 70
Db 63957 GACGGTGTCTAGCACACCGCGGTACACGCGCTGACCGCGCCGCCCGCTCCAG 63898
Qy 71 GluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSerHisSer 90
Db 63897 GTGCGGGAACCCGTGTCGACCGGTCCGCGGTTCGCCACCGCCGACCCCGACGCGCTG 63838
```

```
Qy 91 -----TyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSer 105
Db 63837 AACCTGCGCGCTGAACCTGCCCTCCCTGTACAGCGCGCACGAGGAC----- 63790
Qy 106 LysArgProArgLeuLeuLeuLeuProAspProLeuLeuArgProSerProLeuLeuAla 125
Db 63789 ---CGTACGCGGTTTCAGC---GCCTGAGACGGTCTGCGACACCGCGGCCCC- 63745
Qy 126 ThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrLysLeu 145
Db 63744 -----CGCCCGTA---GGGTGTGCGCCCATGACACACGACCAACACCGTACCGTCCGTC 63692
Qy 146 GluProValSerProProSerProHisThrAspProGluLeuLeuValProPro 165
Db 63691 GAACCCGCGC----- 63683
Qy 166 ArgLeuSerLysGluLeuLeuGlnAsnMetAspArgValAspArg----- 181
Db 63682 -----GTCCGCGAGGAGTGGCCCGCTGCGGACATCGACATCGACACGCGGCC 63629
Qy 182 -----GluIleThrMetValGluGlnGlnIleSerLysLeuLysLys 195
Db 63628 GTCTGCCATCTCGCGAGCGCTTCAAGTGCACCCAGCAGGTGCGGCACCTCAAGGCC 63569
Qy 196 LysGlnGlnLeuLeuGluGluAlaLysProProGluProGluLysProValSer 215
Db 63568 CGCCACCACTGCGCGCCCGCATCGCGCCCGCAGAGGCCAGCA-GATCGCCCGCTCGC 63510
Qy 215 ----- 215
Db 63509 CACCTCGCGGAAAGCGCAAGCTCGACCGCGCTTCCGCGAGAGTTCCTGAATTCAT 63450
Qy 216 -----ProProIleGluSerLysHisArgSerLeuValGlnIleIle 230
Db 63449 CGTCCGCGAGGTATCCACCA-----CGAGCGCATCGCGGA----- 63411
Qy 231 TyrAspGluAsnArgLysLysAlaGluAlaHisArgIleLeuGluGlyPro 250
Db 63410 -----GAGCAACGCGCGCGCGCGCACCGGCCACCGACCGCGCGCCCGC 63357
Qy 251 GlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIle 270
Db 63356 -----CCCCCGGGCGCGACCGCGGGGAGTACGACGCGCG----- 63321
Qy 271 LysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAla 290
Db 63320 -----GAGCGCGCGACCGGACAT 63303
Qy 291 ArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGlu 310
Db 63302 ACCGCGGTACCGAACCGCTCTGTGCTGCCCTGCCAC-----CGCATCAGG 63258
Qy 311 LysLysValGluArgIleGluAsnAsnProArgArgAlaLysGluSerLysValArg 330
Db 63257 CAGCATGTCTGTCATGCTCTACTGACGCGCGACGAAAGCGCA----- 63216
Qy 331 GluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnArgMet 350
Db 63215 -----GACCGGTTCGCGCTCTCTCGACGTCCA---CCACTACCGCATCGACCT 63171
Qy 351 GlnSer-----ArgValGlyGlnArgGlySerGlyLeuSerMet 363
Db 63170 CGATCTACCGCGGGACAGACCTTCGACTCCGACCGCTGA-TCCGCTCAGCGTCC 63112
Qy 364 SerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGlu 383
Db 63111 GCGCGCACCG----- 63100
Qy 384 AsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAsp 403
Db 63099 -----CCGCGACGACACGCTTC-GTCGAGTCAAGCCC----- 63068
Qy 404 GlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyr 423
```


QY 1035 PheAlaGluAlaGlnLysLeuProGlyAspProCysTrpThrSer-----Gly 1052
Db 61269 -----CCGTGCCCGCGGACCCGAACTCGACCGCGAGCTCGGTGGC 61228
QY 1053 LeuProPheProValProProArzGluValIleLysAlaSerProHisAlaProAspPro 1072
Db 61227 GCATCTCACCGCTCGCGTCTCGCGCCACCGACGAGGCGCATCGCGCGGAGC 61168
QY 1073 SerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAla 1092
Db 61167 TGGCGAACGACCGAGCGCCACCGCGGAGGAGTGCCTGCGCGCGCGCTGC 61108
QY 1093 ArgProValLeuProArgPro-----Pro 1100
Db 61107 CGGACCGCGAGCGCAAGCGCCCGCTGGGAGGCGCATGTTCGCTCGGACGACTCTCCA 61048
QY 1101 ThrIleSerAsnProProProLeuIleSerSerAlaLysHisPro-SerValLeuGluAr 1120
Db 61047 ACTACCTGTTCACCGCACCG-----CCAGGGCTTCTGGCAGC 61009
QY 1120 gGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHi 1140
Db 61008 CCGAACAGCGCGAGCTGTCTCGGCGACTACGTACCGCGCTA---CTACCCCGAGCGCGTCG 60952
QY 1140 sAlaLysAlaProValGlyProValThr-----MetGlyLe 1152
Db 60951 CGTCTCGCGCGCG-----CCGCGTCCCGCCATGGCCGACGCGCGCGCGCTGGGCT 60898
QY 1152 uProLeuProMetAspProLysLys---LeuAlaProPheSerGlyValLysGlnGluI 1171
Db 60897 TCCCGCGCCACGCGGTGCGACCGGACACCTCGCGCTGGCGGAGTGCCTCGCGCAGC 60838
QY 1171 nLeuSer-----ProArgGlyGlnAlaGlyProProGluSerLeuGlyValProTh 1188
Db 60837 CCGATCCGATCCCGCGCTCGCGCGCAAGCTCACCGACAGCTCGACGA-----CCTGG 60784
QY 1188 rAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGly-----1201
Db 60783 CGCGCGCTCGCGGTACG-----GGAGCGGCACAGGACTACCGCGCGCGCGCGCG 60730
QY 1202 -SerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspSe 1221
Db 60729 CCCCGCGCGCGCGGCGC-----60711
QY 1221 rAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGl 1241
Db 60710 -----GGCGGTGCCCCCGGTGGGAGCGCGCGGATACCCCTTTCGGGT 60667
QY 1241 yThrIleThrArgIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAs 1261
Db 60666 TGTGATCACTGAGCTGCGCGGGGTACACGCCCCCGCACGTGAAGCTGGAACCCAGCT 60607
QY 1261 pSerLeuProLysGlyHisValIleTyrGluGlyLysGlyHisValLeuSerTyrGl 1281
Db 60606 CCCCGTCCCGGAGGACAGCTCATG-----60582
QY 1281 uGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProPr 1301
Db 60581 -----GGCATGCC-GCCCTCGCTCGGCGCGCGCACGCGCGCGCGCTCGCGCCCTGC 60527
QY 1301 o---HisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyAr 1320
Db 60526 TCGACACCGTGTGCGAGCGCTCGCGCCCGGAC-----GGCGCGACGCGGCG 60479
QY 1320 gAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHi 1340
Db 60478 GCCCGTGC-----CGCGCGCGCGCGCGCGCTCG 60446
QY 1340 sSerProHisLeuLys---GluGlnHisHisIleArgGlySerIleThrGlnGlyI 1359
Db 60445 CCGCCCGGTGGCGGACGCGCGCGCGCCATCTCTGCCCGGAGGAGGCGCGCGCGCG 60386
QY 1359 eProArgSerTyrValGluAlaGlnGluAspTyrLeuArgGluAlaLysLeuLeu 1379
Db 60385 TCCGCGTCTCTGT-----CACCGCGCTGGCGGAGGCGCG 60350
QY 1379 sArgGluGlyThrPro-----ProProProProProSer-----1390
Db 60349 CCGACCGCGCCACCCCTGTGCGCGCGCCACCTGCACACCCCGCGCTCGCGTGGCGG 60290
QY 1391 -ArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLysProAl 1410
Db 60289 CCGCGCCGATCTCGCGGAGCGTCTCAACCCCTC---CCTCGACTCTTGGAGCAGG 60233
QY 1410 ahisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisLysIlePro-- 1429
Db 60232 CACCCCGCGCTC-----CGCGTGAAGCCCTCGTCAACCGCGCTGG 60188
QY 1430 -----ArgGluGluLeuArgHisThrProGluLeuProLeuAlaProAr 1444
Db 60187 CCGGAGACCGCGCGCGCGCGCTGTACCA-CGCGAG-----GCACCGAG 60138
QY 1444 gProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSe 1464
Db 60137 TCCAACAGCTCGCTCTGTCTGCGCGGAGCGGACGCGCGCGCTGGCGGTGTC 60078
QY 1464 rThrThrGlySerLysLysHisAspValArgSer-----LeuIleGlySerProGl 1481
Db 60077 CAGGG-GCGGGCGCCACCATCTGCTGCCCGCGCGCTGCTGCTCGCTGCC-- 60021
QY 1481 yArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluAr 1501
Db 60020 -----GAGCGCTGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCG 59980
QY 1501 gAlaCysTyrGluGluSerLeuLysSerArgProGly-----ThrAl 1515
Db 59979 CGCGCGCTGGAGAGCGCTCACCCAGGTCCCGCGCGCGCGCGCGCGCGCGCGCG 59920
QY 1515 aSerSerSerGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLy 1535
Db 59919 GGGCACCGCGCGCGGCTCATCGCGCTGCCCGAGATCGCGA-CGCTGCGCGCG 59861
QY 1535 sProArgInsProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuPr 1555
Db 59860 CCGACGCGCGCGCT-----CCACGTCGACGCGCGCGCTACG 59825
QY 1555 oArgGlySerProValThrMetArgGluProThrPro---ArgLeuGlnGlySerLe 1574
Db 59824 GCGCGCGCTCTGTTCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 59765
QY 1574 uSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLy 1594
Db 59764 ACACCGTCCCTGGACCTGCACAGCTCGGCTGGCAGCCATCCCGC-----CG 59714
QY 1594 sSerProHisSerThrValProGluHisProHisProIleSerProTyrGluHisLe 1614
Db 59713 GACTCTCTACGCT-----CACCGACGCGCGCGCGCGCGCGCGCGCGCG 59675
QY 1614 uLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspPr 1634
Db 59674 -----TCCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 59646
QY 1634 oThrSerIlePro---ArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProAr 1653
Db 59645 GACGACGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 59598
QY 1653 gHisLeuAlaProAsnProThrTyrProHisLeuTyrProTyrLeuIleArgGlyTy 1673
Db 59597 CGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 59538
QY 1673 rProAspThrAlaAlaLeuGluAsnArgGlnThrIleIle-----AsnAspTyrIleTh 1691
Db 59537 GCGGAGTGGCGCGCGCTGCTGCGAGCGGTGCTCTCCCTGCCCGCGGAGTTCGCGGATCTG 59478
QY 1691 rSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuAr 1711

Qy	2257	tGlySerLysSerProGlyAsnThrSer-----GlnProProAlaaph	2271	REFERENCE	ORGANISM	source
Db	57494	TCGTAGTTGAGTCCGGCGCGCGTTTCGGTGGTTCAGTTTCACGTCACCGCGCA	57435	AUTHORS	Micromonospora griseorubida	
Qy	2271	ePheSerLysLeuThrGluSer-----AsnSer-AlaMetValLysS	2285	TITLE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;	
Db	57434	CCGGCGCGCGTGCACGAGTACTCCGGCGAGTACTCCAGGTACGCCCGCAGCTGCCG	57375	JOURNAL	Micromonosporineae; Micromonosporaceae; Micromonospora.	
Qy	2285	erLysLysGlnGluLeuAsnLysLysLeuAsnThrHis-----AsnArgAsnGluProG	2303	MEDLINE	Anzai, Y., Saito, N., Tanaka, M., Kinoshita, K., Koyama, Y. and Kato, F.	
Db	57374	CCCGTCAGCAGCTTCGCGACCGAGTGTTCGTATACAGTACAGGCTCCACAGTCCCG	57315	PUBMED	Organization of the biosynthetic gene cluster for the polyketide	
Qy	2303	luTyraAsnLysGlnProGlyThrGluLeuPheAsnMetProAlaIleThrGlyThrG	2323	REFERENCE	macrolide mycinamicin in Micromonospora griseorubida	
Db	57314	ATCGTCAGTCGC-CGCGCGGATCTCGGAGTCCGAGGCGCGAGCGCTGTCCGAG	57256	AUTHORS	FEMS Microbiol. Lett. 218 (1), 135-141 (2003)	
Qy	2323	lyLeuMetThrTyraSer-GlnAlaValGln-----GluHisAlaSerThrAsn	2339	TITLE	22472038	
Db	57255	GACCGCAGCAGCGTACTCCGTGCGCGCAGCGCTCTTACACAGCTCTCTCTGCTCCG	57196	JOURNAL	2 (bases 1 to 58343)	
Qy	2340	MetGlyLeuGluAlaIleAlaGlyAlaLeuMetGlyLysTyraSerGlnTrpGluGlu	2359	PUBMED	Anzai, Y., Saito, N. and Kato, F.	
Db	57195	CTTGGTGATCAGTCGATCGGGCGCTCTTGTACCGCGCTCGACCG-TGGTCAGCG	57137	REFERENCE	Direct Submission	
Qy	2360	Ser-----	2360	AUTHORS	Submitted (15-AUG-2002)	
Db	57136	TCTCGTGGCGGTGCCACACCTGTTGACGTAGGCACGACCTTCTCGTCTCGTCCG	57077	TITLE	Pharmaceutical Sciences; 2-2-1 Miyama, Funabashi, Chiba 274-8510,	
Qy	2361	ProProLeuSerAlaAsnAlaPheAsnProLeu-----	2371	JOURNAL	Japan (E-mail: yanzai@phar.toho-u.ac.jp, Tel: 81-47-472-2072,	
Db	57076	CCAGCAGCTCGTATCCCGCGGTCTCTCGACCGGTGGAGTTCGGCAGCAGCCCT	57017	FEATURES	Fax: 81-47-472-2086)	
Qy	2372	-----AsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGly---	2388	Location/Qualifiers	1. 58343	
Db	57016	TACCGACTCGAGTCCAGCGCGCGCTCGAAGACCGAGCTCGAAGCAGGTGA	56957	organism="Micromonospora griseorubida"	2	
Qy	2389	-----ArgSerAspHisThrLeuThrSer-----	2396	/mol_type="genomic DNA"	2	
Db	56956	GCGCTCGCGGAAGCAGCAGCGCTCCGACGACGACGACGCTCTCCCGGTCTCTCGTGA	56897	/db_xref="taxon:28040"	2	
Qy	2397	-----ProGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLys	2415	complement (1..910)	2	
Db	56896	CGACCTTCAGTCGCGGATCTCTCTGCGCGTGCACGACGATCGCGTCGATGCCG	56837	/gene="mycF"	2	
Qy	2416	SerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSer	2435	complement (<1..910)	2	
Db	56836	GCACCTGCC-----GGCCACGTTGGCGCGCGCTCTCGACGTAGGGGA	56792	/gene="mycF"	2	
Qy	2436	GluGlyAspCysAsnArgThrProLeuThr-----AsnArgValTrpGlu	2451	/codon_start=1	2	
Db	56791	CCTGGTCACGTAGGAGGACGTCCCGACGACCGCGAGTGCAGCAGCAGCAGCGT	56732	/transl_table=11	2	
Qy	2452	AspArgProSerSerAlaGlySerThrProPheProTyraAsnProLeuIleMetArgLeu	2471	/product="cytochrome P450"	2	
Db	56731	CCCGGCCATCGAACGAGCTTCGGCACCAATT--TGCGCGCT-----	56690	/protein_id="BAC57023.1"	2	
Qy	2472	GlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeu	2491	/db_xref="GI:28394150"	2	
Db	56689	-----GCTCTCCAGCGCGCGGAAAGCCAGCTTCCTCCAGCAGTACG	56648	/translation="MDRTCAWALPROYAEFRQATCWPAKWGDSPTWLVSRVHVRA	2	
Qy	2492	AlaGlyProHisHis-----AlaTrpAspGluGluProLysPro	2504	LLVDPRVTDTPROPRLSEADGDFRSMMLDPPHTRLRMFISAPSVQVETMR	2	
Db	56647	CCTGTCCAGATCGCGATCCCGGGTGTGTACAGCCCCAGCAGCGCA	56600	PEIEKIVDGLIDELALEPVDILTHALPMSTQVICHLLGVPEDEFFQERSELAS	2	
RESULT 66				RPNDKSMFALLIELVLEGLVTKTAHPDTGLGTAVTERLLKGEITHOELVNNVAL	2	
AB089954/c				LLAAGHETSANQVTLISVLTLRHPTAELRQPELMPNNAVDLLRHSIADGLRRAA	2	
LOCUS				TADIVLDGHTIRAGDGLIILSSANDHNTFGREATFDHRRPARHHVAFGYGPHQCLG	2	
DEFINITION				QNLARLEMEVTLGKLFRRVPAFLRLAQEPDALRVROGSPIFGIDELLVW"	2	
Macromonospora griseorubida gene cluster for the polyketide				2662..2886	2	
macrolide mycinamicin, partial and complete cds.				/gene="mycCII"	2	
AB089954.1 GI:28394148				2662..2886	2	
AB089954				/gene="mycCII"	2	
AB089954.1				/codon_start=1	2	
AB089954.1				/transl_table=11	2	
AB089954.1				/product="ferredxin"	2	
AB089954.1				/protein_id="BAC57024.1"	2	
AB089954.1				/db_xref="GI:28394151"	2	
AB089954.1				/translation="MRIVLDAERCVGAGCCEATAPELFTQDDGLGLVDRPVPVTELL	2	
AB089954.1				GPAREAVDRCPVRAIRIESVVRVGTWARGAG"	2	
AB089954.1				2883..3866	2	
AB089954.1				/gene="mydI"	2	
AB089954.1				2883..3866	2	
AB089954.1				/gene="mydI"	2	
AB089954.1				/codon_start=1	2	
AB089954.1				/transl_table=11	2	
AB089954.1				/product="4-ketoreductase"	2	
AB089954.1				/protein_id="BAC57025.1"	2	

31701 ACCACCGCCAAACGGCGCCCTCACTACACGCGACCGACCCACCACCGCGCACCCCTC 31642
 69 LeuSerGluPheGlnProGlyAsnGluArgSerGlnGluHisArgProGluSer 88
 31641 CTCACCGGACCGCGCCGACACCGCCGCGGATCTCCCTCGGAATGACCCAC 31582
 89 His-----SerTyrLeuProGluLeuGlyLysSerGluMetGluPheLeuSer 105
 31581 CACGCTGGGACACACACCGCCACGACCGCCACCGCCGCGCCACCGCCACCGCCAC 31522
 106 LysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeuLeuAla 125
 31521 AAACGACACGCGGTGCACCATCAACCGCCGCGCCGCGCCGCGCCGCGCCGCGCC 31463
 126 ThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeu 145
 31462 ACCGCTGCGCGCGCA----- 31448
 146 GluProValSerProProSerPro-----ProHis-----ThrAspProGlu 159
 31447 ---CGGCGCAATCCACCAACCGCTGCACACCGCCGCGCCGCGCCGCGCCGCGCC 31391
 160 LeuGluLeuValProProArgLeuSerLysGluLeuLeuGlnAsnMetAspArgVal 179
 31390 ACACCTCGACTCCGCGCA----- 31373
 180 AspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGln 199
 31373 ----- 31373
 200 LeuGluGluGluAlaAlaLysProGluProGluProValSerProProIle 219
 31372 -----ACAAACCGCGAGCATCCCGCCGCGCCGCGCCGCGCCGCGCCGCG 31331
 220 GluSer-LysHisArgSerLeuValGlnIleIleTyrAspLeuAsnArgLysLysAlaG1 239
 31330 GAAACACAAACAC----- 31292
 239 uAlaAlaHisArgLysLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnG1 259
 31291 AACCCACACCGCCGAGAACCCACACCGCCGCGCGCGCGCGCGCGCGCGCGCGCG 31244
 259 nProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLy 279
 31243 ACCCAACCGCCGAGAACCGACGACCGCCGCGCCGCGCGCGCGCGCGCGCGCGCG 31184
 279 sLeuIleLeuTyrPheLysArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCy 299
 31183 A---CCACCCACATCCACCGCGAACAGACGACGACGACGACGACGACGACGACGAC 31132
 299 sGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgLysGluAsnAs 319
 31131 -----AAGCGGACCGAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 31106
 319 nProArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProG1 339
 31105 AAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 31067
 339 uIleArgLysGlnArgGluGlnGluArgMetGlnSerArgValGlyGln-ArgGlyLys 359
 31066 CCGACAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 31007
 359 eGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleLeuAspGlyL 379
 31006 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30947
 379 euSerGlnGlnGlnAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetL 399
 30946 TGCGGAAAGAGAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30903
 399 euTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAsp 419
 30902 -----GGCCACGACGCGCTCC 30884

419 roMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluT 439
 30883 CCAC-CACCTCCACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30834
 439 hrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuG 459
 30833 CCCACATGCAACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30798
 459 luArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnT 479
 30797 ATCGCCAA----- 30762
 479 yrLysSerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGln----- 495
 30761 GCAGCAGCTGAGCATGACCAATATTCGACTTCACCGACCGCCCAACAACGCGCAACCA 30702
 496 -----GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 508
 30701 CCACCGACGCGCGCATCGTCGCAACACCGCTCGCGCTCAATCGGATCCCGCCACCGC 30642
 508 lngGlnProMet-ProArgSerSerGlnGlnGluLysAspGluLysGluLysGluLys 527
 30641 GTCCCGCTCCATGCGCTCCCGCCATCCACCATCCACCATCCACCGACGCGCGCGCGCG 30582
 528 GluAlaGluLysGluGluGluProGluValGluAsnAspLysGluAspLeuLys 547
 30581 AACGCGCGCGCA----- 30564
 548 GluLysThrAspAspThrSerGly-----GluAspAsnAspGluLysGluAlaValAla 565
 30563 CGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30504
 566 SerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgile----- 582
 30503 GCGGAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30450
 583 ThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAla 602
 30449 AACCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30390
 603 GluLeuAlaSerMetGluLeuAsn-----GluSerSerArgTyrThrGluGluMetGlu 621
 30389 GAGAACGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30331
 622 ThrAlaLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetVal 641
 30330 GACATCCGCGCGTGGC-----CATCACCGT----- 30307
 642 GlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsn 661
 30306 -----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30262
 662 LeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArg 681
 30261 CTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30202
 682 LysLys-----LysLysAla 686
 30201 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30142
 687 ProAlaAlaAlaSerGluGluAlaAlaPheProValValGluAspGluGluMetGlu 706
 30141 GTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30082
 707 AlaSerGlyValSerGlyAsnGluGluGluMetValGluAlaGluAlaLeuHisAla 726
 30081 CCCATTGTGGTCCCAT-----GTAGACCGCGCGCGCGCGCGCGCGCGCGCGCG 30041
 727 Ser-----GlyAsnGluValProArgGlyGluCysSerGlyProAlaThrVal 742
 30040 GCTCAACGCGCGGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29996

Db	28196	TCGGCGATGTCCCGTCCGTAGC-----AGGGAGCGATCGGCG	28158
Qy	1366	laGlnGluAspTyrLeuArgArgGluAlaLysLeuLysArgGluGlyThrProPro	1386
Db	28157	AGCAACAGGACCCAGCGTCCGTGAGCGGCGC-----	28126
Qy	1386	roProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuL	1406
Db	28125	-----GTCCTCGTTCC	28116
Qy	1406	ysLeuLysProAlaHisGluGlyLeuValAlaThrVal-----	1418
Db	28115	ATCGGGACCGGTGCCAACGTACGGGTACGGAGGTCTCTGCTGTCGTGGGGGTG	28056
Qy	1419	-----LysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrP	1437
Db	28055	GTCCCGCGGTGTGGGGGAGCCAGTACCGGTCTGTGTAACGATACGTGCGCAAC	27996
Qy	1437	roGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProL	1457
Db	27995	CCCACTGACGCCACACCCCAAGAACAGAACCCCGCGCCACTGCACCCGGCAACCCCG-	27937
Qy	1457	eulysTyrAspThrGlyAlaSerThrGlySerLysLysHisAspValArgSerLeuI	1477
Db	27936	-----CACCAACCTCAGCGCACTCACAC-----	27910
Qy	1477	legIysProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAla	1497
Db	27909	--CATCGGCCCAACACCATCCCGCCGACG-----	27880
Qy	1497	rgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerS	1517
Db	27879	-----CAAGAACCCACACACACACCCACCC-----	27850
Qy	1517	erSerGlyGlySerIleAlaArgGlyAlaProValIleValProGlu---LeuGlyLysP	1536
Db	27849	-----ACCCGCCACCTCCCGCCACACATCTCAATCGCC	27816
Qy	1536	ro-----ArgGlnSerProLeuThrTyrGluA	1545
Db	27815	CCCAACACCGGATGCGCGACACCTCCACAAACGACGAAACCCCATCACCACCAAC	27756
Qy	1545	spHisGlyAlaProPheAlaGlyHisLeuProArgGlySerPro---ValThrMetArg	1564
Db	27755	CG-----ACGCACACCTCTCCCAAAACGACCCCGCAACCCCAATTCTTA	27711
Qy	1564	luProThrProArgLeuGlnGlySerLeuSerSerSerLysAlaSerGlnAspArgL	1584
Db	27710	AACCAATACCCCATCCCAAAACCCACCATCCACACACCCACAT-CCACAGACGAATA	27652
Qy	1584	ysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHis-	1603
Db	27651	AA--ACGGAAATCACCGACTCAC-----CGCACGAACCTCCCGCAACACACT	27607
Qy	1604	-----HisProHisProIleSerProTyrGluHisLeuArgGlyValSerGlyV	1621
Db	27606	GAGCAACTATCCCGCACACATCCACCAACCCCAAT-----	27569
Qy	1621	alAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyI	1641
Db	27568	-----GCGACCATATACCCACCGCAACCAACCGCGCGCGCAACCCCA-----	27527
Qy	1641	leProLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsn-----P	1659
Db	27526	-----CTTCCCAACACCGCACCAACCAACCCCTC	27499
Qy	1659	roThrTyrProHisLeuTyrProPro-TyrLeuIleArgGlyTyrProAspThrAlaAla	1678
Db	27498	CACCCACCCCATCCTCCCGCCACACACCATCCTGCGCCG-----ACC	27457
Qy	1679	LeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisAsn	1698
Db	27456	ATTAACCAACCGCCACTCCAAACCAACCCCGCCAAACCCCGCCAAACCCCTCCACCAACC	27397
Qy	1699	ThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSer	1718
Db	27396	CACCGACCCCGCCACCGACCATCGCCCGCCACCGCCCA-ACCGCGCCCA-----	27344
Qy	1719	SerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValPro	1738
Db	27343	-----TCACCGCGACCGACCCACCA-----	27323
Qy	1739	HisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeu	1758
Db	27322	-----CCACCGCGCACCTCTCCACCGCAACGCCC-----	27290
Qy	1759	AlaTyrLeuProThrAlaProGlnProPheSerArgHisSerSerSerProLeu---	1777
Db	27289	-----CCGACACACCGCGCGCGGATCTCCCTCGAATGACCCACCCCGCT	27239
Qy	1778	SerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGlu	1797
Db	27238	GCGGCACACACCCCGCAACCGCAACACCCCGCGCAACCGCCATCACCACAA-----	27185
Qy	1798	ArgAspArgAspArgGluArgAspArgAspArgGluArgLysSerIleLeuThrSer	1817
Db	27184	-----ACG	27182
Qy	1818	ThrThrThrValGluHisAlaProIleTyrArgProGlyThrGluGlnSerSerGlySer	1837
Db	27181	ACACCGCTGCACCAT-----CAACCGCGCCCAATGACCCCGCCCGACT	27134
Qy	1838	SerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAla	1857
Db	27133	CACCAACCGCTCGCGCGCACCGCCCAATCCCAACACCCCTGCACCCCGCCCAACCT	27074
Qy	1858	HisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerVal	1877
Db	27073	CATCAACCGCTCCGAAACACCTCGGACTCGCCA-----	27038
Qy	1878	LeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrVal	1897
Db	27037	-----ACAAACCGCGCGCATCCCGCCCGCCACTGCG	27008
Qy	1898	LeuArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProAlaThr	1917
Db	27007	CCCGCTGACCGGAAACACAAACACACACACCGCCGAC-GCCACACCGCCCGCAACCCACC	26949
Qy	1918	HisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeu	1937
Db	26948	ACCACCCCA-----GAACCCACACCA	26928
Qy	1938	LeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspThr---Gly	1956
Db	26927	CCCGCGCAACACACAC-----AACCCACCCACACCCCGCAACCGCCCGCCCGC	26877
Qy	1957	HisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSer	1976
Db	26876	CACCCAAACACCGAAGACACCC-----AACGACACCCCGCATCCACACCCGAA	26826
Qy	1977	LysGlySerGluProArgProLeuValProValSerGly-----	1990
Db	26825	ACAGCACCAACCCAGAAC-----GAACCGGAAAGCGGACGAGAAAGCGCGCTCC	26772
Qy	1991	-----HisAla-ThrIleAlaArgThrProAla-----	1999
Db	26771	AAACGACCGCGCGCGCAAAACCCCTCCGACCGCGCCGACCAACCCCGGCACA	26712
Qy	2000	-----LysAsnLeuAlaProHisHisAlaSerProAspPro-----	2011
Db	26711	GCAACAGAGAGAACCGCGTCCGACACATCCGACACCGGACCCACCCACAGAGAA	26652
Qy	2012	-ProAlaProProAlaSerAlaSerAspProHisArgGlyLysThrGlnSerLysProPh	2031
Db	26651	TCGGCGCGCGCAACAGACACCGGAAACCAACAAAGAGCGGCATCCGCAACCCGGA	26592

```

Qy 2031 eSerIleGlnLeuLeuAlaArgSerLeuGlyTyrHis----- 2044
Db 26591 GCCTGCTCCAAATACATCGCGCGTTCG-----TACCACTAAACCCGGAACGACGACACC 26538
Qy 2045 -----GlySerSerTyrSerProGluGlyValGluProVa 2056
Db 26537 GCCCGCCGACCGCGCGACCGTTCGCGGCACGACGACGCTCCCGCACCATCTCCACC 26478
Qy 2056 lSerProValSerProSerLeuThrHisAspGlyLeuProLysHisLeuGluG1 2076
Db 26477 GCACCCGACGACCAATCCA-----CATCGGAAGACGCGCACACCCACAT----- 26435
Qy 2076 uLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLe 2096
Db 26434 -----GCAACGAAGCGCGCACCA----- 26417
Qy 2096 uGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuGlnProSerSe 2116
Db 26416 -----CCCCAC-----GCTCCATCGGCAACACCATCTTC 26388
Qy 2116 rSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAl 2136
Db 26387 ACCACCC-----CAGCCACCCCGCAGCAGCCTGAGATGACCAATATTCGACT----- 26339
Qy 2136 aGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSe 2156
Db 26338 -----TCACCGACCCCAACAACAAAGACGACCAACCCACGACACCCATACGTC 26289
Qy 2156 rAlaProLeuProAlaLeuTyrSerPheProGlyAlaSer-----CysProVa 2173
Db 26288 GCCAACACCGCTCGCGCTCAATCGGATCCCGCCACCGCGTCCCGCTCCATGCCCTCC 26229
Qy 2173 lLeuAspLeuArgProProSerAspLeuTyrLeuProProPro----- 2188
Db 26228 ACCATCCCATCATCACCCACGACCAAGCGGCACGACGCGCGCGGCAATCACCCGC 26169
Qy 2189 -----AspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysAr 2205
Db 26168 TGCTGCGACGACGACCTCGGGCGCGTCAAAACCATTCGACGACCATCTGATTAAACGCC 26109
Qy 2205 gSerProGluProAsnLysThrSerValLeuGlyGlyGluAspGlylleGluProVa 2225
Db 26108 GAACCCCGCACCCCGCCAAACA-----CCCGACGACGACGACGACCC 26067
Qy 2225 lSerPro-----ProGluGlyMetThrGluProGlyHisSe 2237
Db 26066 GCATCCGACACCGCTGCACCAACACACACCCACGCTCGGACCAACCGTACCATCA 26007
Qy 2237 rArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMe 2257
Db 26006 GCCGAATCCGAATACGCT-----TACACCGAC-----CATCCACGCGCAACCCACCC 25959
Qy 2257 tGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSer-LysLeuThrG 2277
Db 25958 TGCGCGCAAACTCCA-----CAAAAATCCCGCGCGTGGCCATCACCGTCACCCCA 25908
Qy 2277 luserAsnSerAlaMetValLysSerLysGlnGluIleAsnLysLysLeuAsnThrH 2297
Db 25907 CCAGCCAAACGCCATCATCACTGCCAGACGCGCAACGCTGACCGCCGATGCAACGCC 25848
Qy 2297 isAsnArg----- 2299
Db 25847 ACCAACGACGACGACGACGCGTGTCCACCGTCACCGCGCGCCCTCCAAACCAACACA 25788
Qy 2300 --AsnGluProGlyTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaI 2319
Db 25787 TAAGACACCCGACGACGACGACGACGCTCCCGCGCGTGGCGGAGAAAGCCCGCGTC 25728
Qy 2319 lThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisLeuSerThrA 2339
Db 25727 TCCTGTCGAGGAGGATGGTT-----CAGCCGCTAGTCGTGGTCATCACCGCGCGCTAG 25674
Qy 2339 snMetGlyLeuGluAlaIleAlaLeuMetGlyLysTyrAspGlnTrpGluG 2359

```

```

Db 25673 ACACCGGTACGGTCCCGCGCAACGC----- 25639
Qy 2359 luserProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeu--P 2378
Db 25638 AAAACCCGCGCGTCCAAACGCTCCCGACGACCTCCAAACAAACGCTCTCGGATC 25579
Qy 2378 roAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProG 2398
Db 25578 CATCGCCACACGCTCATCGCGCGGACACACCAAAAAACCCCGCATCG-----AAACCCG 25526
Qy 2398 lGlyGlyGlyAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProA 2418
Db 25525 GCGCATCGACACG-----AACCCACCGCTCCCGCAATACATCGCGCGCG 25481
Qy 2418 la---ProGlyLeuAlaSerGlyAspArgProProSerValSer-SerValHisSerGlu 2436
Db 25480 CACGCGCCGCG-----TCGGGATCGAAGATGCGGTGCGAGTCCGAGTGTGGTCCGTCG 25427
Qy 2437 GlyAsp-----CysAsnArgArgThrProLeuThrAsnArgValTrp 2450
Db 25426 GGAACGGGAGATCGCGTCCGGGAGGAGACAGCTCCCACTGTTCGGGCGGTTT 25367
Qy 2451 GluAspArgProSerSerAlaGlySerThrProPheProTyrAsn---ProLeuIleMet 2469
Db 25366 CCACCCACCGGAAAC-----GGCAGCCCATTCACGATCGCCACCGCGTGTGTG 25313
Qy 2470 ArgLeuGlnAlaGlyValMetAlaSerProProGlyLeuProAlaGlySerGly 2489
Db 25312 CCTCTTCGGTCAACCGGTGCACCTGTGCGCTCGTCCGAGGTCCGCCAGGCGACGCG 25253
Qy 2490 ProLeuAlaGlyProHisHisAla-----Tyr 2498
Db 25252 -----GAAGTACTACGACGCTTCTTCGTGTGTGGCAGTCACGCGCTTCCCTGG 25199
Qy 2499 AspGluGlu 2501
Db 25198 GATCGTCAG 25190

```

RESULT 67

AC026106

LOCUS

DEFINITION

AC026106

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 164485)

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-oshman, P.R., Allen, C.,

Alsbrooks, S.L., Amarantunga, H.C., Are, J.R., Banks, T., Barbara, J.,

Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J.,

Bowen, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,

Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F.,

Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,

Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,

Coley, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,

Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,

Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,

Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,

Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,

Foster, P., Frantz, P., Gabis, A., Gao, J., Garcia, A., Garner, T.,

Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,

Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,

Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,

Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,

Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,

Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korwar, J.,

Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,

Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,


```
Qy 1466 rGlySerLysHisAspValArgSerLeuLeuLeuGlySerProGlyArgThrPheProPr 1486
Db 138744 ACCATCCAGCCAGCCAGCCAGTCAGCCAGCCGAGCAGCCAGCCAGCCAGCCAGCC 138803
Qy 1486 ovalHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCys----- 1503
Db 138804 AGCCAAAGCC-----GCCAACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 138845
Qy 1504 -----TyrGluGluSerLeuLysSerArgProGlyThrAlaSerSe 1517
Db 138846 CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 138905
Qy 1517 rSerGlyGlySerIleAlaArg-----GlyAlaProVal-- 1528
Db 138906 CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 138965
Qy 1529 -IleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAl 1548
Db 138966 GCATCCAGCCAACTAAGCCAGCCG---GCTAGCCAAAGCCAGCCAGCCAGCCAGCCAGCC 139022
Qy 1548 aProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrProAr 1568
Db 139023 CAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 139076
Qy 1568 gLeuGlnGluGlySerLeuSerSerLysAlaSerGlnAspArgLysLeuThrSerTh 1588
Db 139077 GCCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 139124
Qy 1588 rProArgGluLeuAlaLysSerProHisSerThrValProGluHisHisProHisProIl 1608
Db 139125 GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 139164
Qy 1608 eSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIl 1628
Db 139164 ----- 139164
Qy 1628 eProLeuAlaPheAspProThrSerIlePro-----ArgGlyIleProLeuAspAl 1645
Db 139165 -----CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 139208
Qy 1645 aAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrPro----- 1662
Db 139209 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 139268
Qy 1663 -----HisLeuTyrProTyrLeuLeuLeuArgGlyTyrProAspThrAl 1677
Db 139269 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 139325
Qy 1677 aAlaLeuGluAsnArgGlnThrIleLeuAsnAspTyrIleThrSerGlnGlnMetHisHi 1697
Db 139326 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 139361
Qy 1697 sAsnThrAlaThrAlaMet-----AlaGl 1705
Db 139362 ACCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 139421
Qy 1705 nArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAl 1725
Db 139422 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 139481
Qy 1725 aAlaGlyProArgGlyIleLeuAspLeuSerGlnValProHisLeuProValLeuValPr 1745
Db 139482 AAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 139535
Qy 1745 oProThr-----ProGlyThrPr 1751
Db 139536 GCCAACTTAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 139595
Qy 1751 oAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaPro----- 1765
Db 139596 ACCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 139655
```

```
1766 -----GlnProPheSerSer----- 1770
139656 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 139715
1771 -----ArgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysPr 1788
139716 GGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 139775
1788 oThrThrThrSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspAr 1808
139776 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 139791
1808 gGluArgGluLysSerIleLeuThrSerThrThrValGluHisAlaProIleTPrAr 1828
139792 -----CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 139811
1828 gProGlyThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySer-- 1847
139812 GCCTCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 139871
1848 -SerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProIleArgTh 1867
139872 GCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 139931
1867 rGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIl 1887
139932 CCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 139988
1887 eThrAlaValGluProSerLysProThr-----ValLeuAr 1899
139989 GCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 140048
1899 gSerThrSerThrSerSerProValArgProAla----- 1910
140049 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 140108
1911 -----AlaThrPheProAlaThrHisCysProLeuGlyGlyTh 1924
140109 AAGCCATCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 140163
1924 rLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProAr 1944
140164 -----CATCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 140209
1944 gValAlaArgProGluArgProArgAlaAspThrGly-----HisAlaPhe----- 1959
140210 GCCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 140269
1960 -----LeuAlaLysPro-----ProAlaArgSerGlyLeu-----GluProAlaSe 1973
140270 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 140329
1973 rSerProSerLysGlySerGluProArg----- 1982
140330 GCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 140389
1983 -----ProLeuValProProValSerGlyHisAlaThrIleAlaArgThrPr 1998
140390 AAGCCATCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 140446
1998 oAlaLysAsnLeu-----AlaProHisHisAlaSerPro-----AspPr 2011
140447 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 140506
2011 oProAlaProProAlaSerAlaSerAspProHisArg---GluLysThrGlnSerLysPr 2030
140507 ACCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 140566
2030 oPheSer-----IleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGl 2045
140567 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 140626
2045 ySerSerTyrSerProGluGlyValGluProValSerProValSerSerProSerLeuTh 2065
```


Db 10212 ATCCACACCGCCGACCTCCAAACACCCGACCCGCAACACCGGATGCCCGGAA-----CA 10156
Qy 89 sSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProAr 109
Db 10155 CTCCACAAACCCGAA----- 10140
Qy 109 gLeuGluLeuLeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnPr 129
Db 10139 -----AACCATCCCCA-----ACAAAGCCGCC 10117
Qy 129 oAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluPro-Vals 149
Db 10116 CACCACATCCGAAACAAACACCCGCTCAGCAGATTCGGATACCAATACCCACCATCAA 10057
Qy 149 erProSerProProHisThrAspProGluLeuGluLeuValProArgLeuSerL 169
Db 10056 ACCACCCCATCAACACACACCCACCCGCGAGATAAAACGGACACACCCCGCCC 9997
Qy 169 ysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGln 189
Db 9996 CGGAC----- 9992
Qy 189 lnIleSerLysLeuLysLysGlnGlnGlnLeuGluGluGluAlaAlaLysProProG 209
Db 9991 -----ACAAACCCCAACACCCGACAGCATCCCTCCACCG 9952
Qy 209 luProGluLysPro-----ValSerProProIleGluSer-LysHisArgSer 225
Db 9951 CTCACATGCGCGGATGCGAGCATATATCCACCGCACCGCGCGCACCCACAC- 9998
Qy 226 LeuValGlnIleTyrAspGluAsnArgLysLysAlaGluAlaHisArgIleLeu 245
Db 9898 ----- 9898
Qy 246 GluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGln 265
Db 9897 -----CCATCCCGCTCACATCCACCAACAAACCC----- 9868
Qy 266 TyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLys 285
Db 9867 ---CACAGCAGATCCATCCCGCGACACACCGTCACGCGGACCATTCAC---CGC 9814
Qy 286 ArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeu 305
Db 9813 CGGCACACCAACCGACACCCACCCCGCCGCAACCGTTCCTGACATCCACGATCGGCAG 9754
Qy 306 MetGluAlaLeuGluLysLysVal-----GluArgIleGluAsnAsnProArgArg 323
Db 9753 TCCGACCGACACCATCGACACACCCGCGCAACACCTCACCGCATCCCGCGCA----- 9700
Qy 324 AlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGln 343
Db 9699 -----CGCAACCGCCACCCCG 9682
Qy 344 ArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySer---GlyLeuSer 362
Db 9681 CGCCCCATCAGCACACTCAACACACACCCCGCGCGAGCAACCTCACCTCGCA 9622
Qy 363 MetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGln 382
Db 9621 ATGACCCACACCGCATCCCGGACACACACCCCGCA-----CATCCACAAACCGCGCAA 9568
Qy 383 GluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAla 402
Db 9567 CGACCATCAGCCCGCCGACACCGCGCTGACCATCAACCGGACCCACCCCGCCCTC 9508
Qy 403 AspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysVal 422
Db 9507 ACCCGCCCGCGCGCACCTCCAAACAA-----ATCCCAATC 9469
Qy 423 TyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGlu 442
Db 9468 CACAAACCCCGA-----CAACGCGCGGACCACTCCACCAT 9433

Qy 443 LysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThr 462
Db 9432 CGACTCGCGCAACAC-----CTCCGACTCGTCCAGCAGCGCGCAC 9391
Qy 463 ValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeu 482
Db 9390 CATACCCACCCACTG-----CCAACCTGACCGGAAACACAAACACCCCGCAC 9337
Qy 483 Val-----ArgArgSerTyrArgArgGlyLysSerGlnGlnGln 496
Db 9336 ACCACAGCAACACCCGACCGCACCGAACCCACACACCGAACCCGAGCCCGGCC- 9278
Qy 497 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetPro----- 513
Db 9277 CAGAGCCACCGCACCAAGACAGCAACAAACC-TCTCCACACCCCGCACCGACGACC 9219
Qy 514 ---ArgSerSerGlnGluLysAspGluLysGluLysGluLysGluAlaGluLysGlu 532
Db 9218 GCCGATGCTTCAACACCGACCGCGACGACCAACACCAACCCACATCCACCAACCA 9159
Qy 533 GluGluLysProGluValGluAsnAspLysGluAspLeuLysGluLysThrAspAsp 552
Db 9158 GAACACCCCGGAG-----CCACACCGCTCAGAC 9129
Qy 553 ThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThrAla 572
Db 9128 AACCCCGCGCACCTCACGCAACCGACCGCATCGCC-----CGCAAGCCCCC 9078
Qy 573 AsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSer 592
Db 9077 ACCGAACCCCGCACACACCCGCAACCGCAAGCTCTCTGCCAACACCAACCCCTCG 9018
Qy 593 GluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSer 612
Db 9017 CCGCCATCCAGACACCGCATCGGAATCGGA----- 8985
Qy 613 SerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArg 632
Db 8984 -----TCGGAATCGGAATCGGCA----- 8961
Qy 633 AsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsn--- 651
Db 8960 -----GGATCGTCGAACCTCGGGGCGCTGCTCCAAATC 8928
Qy 652 -----PheTyrPheAsnTyrLysArgGlnAsnLeuAspGluIleLeuGlnGln 668
Db 8927 AGATGGCATTCGTCC- CGACACCCCAACCGCGACACACCCCGCACGACGACGAC 8869
Qy 669 HisLysLeuLysMetGluLysGluArgAsn----- 678
Db 8868 CACACCGGCCAATCCCGCGCTCCGTCACCAACTCCACCCCGGAGACCAATCCAC 8809
Qy 679 -----AlaArgArgLysLysLysAlaProAlaAlaAla 690
Db 8808 CAACGCGCGACGACATCCACATGCAACGTGCGCGCA-CCACCCCGCGCGACGCA 8750
Qy 691 SerGluGluAlaAlaPheProValValGluAspGluMetGluAlaSerGlyVal 710
Db 8749 TCACCATTTGATCACACCGCGCACACCCGCGCGCTGCGTATGACCAATATTCGACT 8690
Qy 711 SerGlyAsnGluGluMetValGluGluAlaLeuHisAlaSerGlyAsnGlu 730
Db 8689 TCACCGAACCAACACACCGCCCGCACGACCGCCCGCCATAGC----- 8639
Qy 731 ValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAsp-----Thr 748
Db 8638 -----CCAACACCGCTCAGCTCAATCGGATCCGATCCCAAC 8603
Qy 749 GluSer-----IleProSerProHisThrGluAlaLysAspThrGlyGlnAsn 765
Db 8602 GAGTCCCGCTCCATCGCGCTCCACCATCCACAT-----CCGACACACCCCAAC 8552

766 GlyProLysProProAlaThrLeuGlyAlaAspGlyProProGlyProProThrPro 785
8551 CGCCGCGCACCAACGCGCGCACCAATCACCGCGCTGCGCCA-----CCCGCGAGCGG 8498
786 ProArgArgThrSerArg-----AlaProIleGluPro 796
8497 CCCTCAAAACATTGACCCACCATCTGATTCAACCGCAGAACCCCGCACCA---CGGCCA 8441
797 ThrProAlaSerGluAlaThrGlyAlaProThrProAlaProProProProSer 816
8440 ACACCCGATACCAACGCTCCCGCGCCAGCAGAACCGCTCAACACCAACACCCCGCAC 8381
817 AlaProPro---ProValValProLysGluGluLysGluGluThrAlaAlaPro 835
8380 CCTCCCGCCACCCCGCTCCATCCGACCAACCCGCAACCGCCGACACCCGCAATCACCG 8321
836 ProValGluGluGlyGluGlu-----GlnLysProProAlaAlaGlu----- 850
8320 ACAACCCACCTCGCGAGAACTCCACAAACACACCCCGCGCGCAATCACCGTCAACC 8261
851 -----LeuAlaValAspThrGlyLysAlaGlu 859
8260 CACCGCGCAACGCGACATCACCTACCCAAACGCAACCCCTGACACGCGCAATGCAACG 8201
860 GluProValLysSerGluCysThrGluGlu----- 869
8200 CCACCAACGACGACGACACCGCGTATCAACCGGAATCGACGACCCCTCCACCCCAACA 8141
870 ---AlaGluGluGlyProAlaLysGlyLysAsp-----AlaGluAlaAlaGlu 884
8140 CATACGAAACACGACCCGCGCGAGCTCACCGACTGCGCGTGAGCAGATAGCCCTCGA 8081
885 AlaThrAlaGluGlyAlaLeuLysAlaGluLysGluGlySerGlyArgAlaThr 904
8080 GTTCCTCGGAGCTCTGTCGAGACGGCGGCATCTCTGCTGGGTGAGCGCGAGTACA 8021
905 ThrAlaLysSerSerGlyAlaProGlnAspSerSerAlaThrCysSerAlaAsp 924
8020 ---CGCGGTCGCGTACCCGCGAC-----CCGCGCGAT 7988
925 GluValAspGluAlaGluGlyGlyAspLysAsnArgLeuSerProArgProSerLeu 944
7987 CGATTCCGCGACGCT-----CCACGCGCTCCACG 7958
945 LeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLys 964
7957 ACGTCTCCAAACAACGAGCGCTGCTGCGGATCCATCGCCACCGCTCACGCGGCGACAC 7898
965 Gln-----LeuLysGlnArgAlaAlaAlaIleProProIleGlnVal 978
7897 CAAGAAGCTCGATCGACATCCCGGTATACACAAACCCCTCCCGCGCATACG 7838
979 ThrLysValHisGluProProArgGluAspAlaAlaProThrLysProAlaPro----- 996
7837 ACTTCCCGCGCGCGCGCGAGGAGTCAAGAGAGGAGTCCACGCTCCAGCCCGGTCGT 7778
997 -----ProAlaProProProGlnAsn----- 1004
7777 CGGGGCTGGGCGGATGATCGTCAGCACCGCGCAGCAGGAGTCCAGAGTCTTCGGGAG 7718
1005 -----LeuGlnProGluSerAspAlaPro----- 1012
7717 TGTGGCTCCCGGGGTAGCGGAGGCCATGCGCAGCATGCAATAGGCTCGTCGACGG 7658
1013 -----GlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaPro 1028
7657 GCACGCGTTCGCGAGTCAACCGGA---CTTCTCGGCAAGACGCGCGACGAGCGCCT 7601
1029 ProAlaAspLysGluAlaPheAla-----AlaGluAlaGln 1040
7600 CAGGAAGTCTGCCAACTGGACGGGTGGGTGTCGAAGACCAACCGTCGAGGCGAC 7541
1041 Lys----- 1041

7540 TCAGACCGGTCTGCTGCTGAGTGTTCGAGTTCGAGCGCGCTGAGCGAGTCGAAGC 7481
1042 -----LeuProGly-----AspProProCysTriThrSerGlyLeu 1053
7480 CGCGCTCTTGAAGCCCGGGTGGCGCGACCGCATCCACCGTCTCTGTCCTCAGACCG 7421
1054 ProPheProValProProArgGluValIleLysAlaSerPro-HisAlaProAspProSe 1073
7420 CGCGCGCGCGCGCC-----GCACGTTTCGAGGAGGACCGCT 7382
1073 rAlaPheSerTyAlaProProGlyHisProLeu-----ProLeuG1 1087
7381 GCGTTTCGCGCTCCCGCATCCCGCCACCGCTGCGCAAGCGCGCTCGCGCGACTCGG 7322
1087 yLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnPro----- 1105
7321 TGTCCAC-----TTTCGGGATCGAGCGCTCTCTGGCCTCGGCGACCCCGTCA 7271
1106 -----ProProLeuIleSerSerAlaLys----- 1113
7270 AGAGCGCGCGCGAGCGCTCGCAAGACCGCGGAGCGAAACCGTCCAGTCGACGCTCG 7211
1114 -----HisProSerValLeuGlu----- 1119
7210 CGACGCTGACACACACGTCATCGCGTTCGAGCGCTCTGTCAGAGCGCGATGCCGAGCT 7151
1120 -ArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTySerG1 1139
7150 CGGGTTCGAAGCGCGTATGCCACGCGGCTTTCAGGTTTCGCTCGCGGATCCCTCGCCA 7091
1139 uHisAlaLysAlaProValGlyPro----- 1147
7090 TGCCACCGCGCGCCACAGTCCCAAGTATCGAGTGGCGGGAAGACCTCGGCGACGAC 7031
1148 ----ValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerG1 1166
7030 GTCGCTCCCGCAGGCGCT-CCAGCGATGCTTTCGCGCGCGCTACCGCCCT-----GG 6978
1166 yValLysGlnGluGlnLeu-----SerProArgGlyGlnAlaGlyProProGluSe 1183
6977 CGCGCTTCGCCAGACTCCGTAACCGATGAGAGAGGACGAGCGCTCCAGTCTCTCG 6918
1183 rLeuGly----ValProThrAla-GlnGluAlaSerValLeuArgGlyThrAlaLeuGlyS 1202
6917 TGTGGTGTGAGTTCGTCGAGCAGCAGCGCGCGAGCGTCTTCGAGCGAAGACTCTCGTCG 6858
1202 erValProGlyGly-SerIleThrLysGlyIleProSerThrArgValPro----- 1218
6857 ATGTGTCGCGCGCGAGCTGTCGATC---ACACGCTGTCGAGGACTCCGGCGGTGTGG 6801
1219 -----SerAspSerAla-----IleThr 1224
6800 AAGACCGAGTGGCGGATGCGCGCGCAGGACGTCGAGCAGCGCGCTCCGCTACA 6741
1225 TyrArgLysSerIleThrHisGlyThrProAlaAspValLeuTyLysGlyThrIleThr 1244
6740 TCGAGCGCGCATGTCACTCAACACCGAGC---GCGCAAGATCAGTCTCCAGTTCT 6684
1245 ArgIleIleGlyLysAspSerProSerArgLeuAspArgGly----- 1258
6683 GCGCGCGCGCGGCTCGGCACCCCGACGACTGTGAGAACAGAGGTGTCGCGCACCGTTG 6624
1259 ArgGluAsp-SerLeuProLysGlyHisValIleTyGluGlyLysGlyHisValle 1278
6623 CGAGCGAGCACCGCGGCCACCCCGCCACCTCCCGTACTCCCGTCCCGTCCAGGATG 6564
1278 uSerTyGluGlyClyMetSerValThrGlnCys-----SerLysGluAspG1 1294
6563 TTCCGCGAGCGGCCATTCCTCCGCGCGCATCCCGCGCGAGTCCCGCGCGGTCACCCGCGCG 6504
1294 yArgSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyAspMetMe 1314

Db 6503 TAGATCCCGGAGCGCGGTACCGGA-----TCTGGTCTCTCCCGCTCGCTC----- 6458
Qy 1314 tGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAl 1334
Db 6457 -----CGGCGAGGGCGT-----TGCGGAGC 6438
Qy 1334 aileProProGluArgHisSerProHisHisLeuLysGluGlnHisIleArgGlySe 1354
Db 6437 TGGTCCAGCATCCGGCGATTCC-----GTACGG----- 6410
Qy 1354 rIleThrGlnGlyIleProArgSerTyrValGluAlaGlnIleAspTyrLeuArgArgL 1374
Db 6409 -----ACGGAAGT----- 6401
Qy 1374 uAlaLysLeuLeuLysArgGluGlyThrProProPro-----ProPr 1389
Db 6400 -----CGACGAGCGCGCCCGCGGTCCGGCAGTTCCAGGGCGGCCACC 6357
Qy 1389 oSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysPr 1409
Db 6356 CGCCCGAAGCGCCACACTCGCGCACCGAACACAT-----CC 6321
Qy 1409 oAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGlu----- 1427
Db 6320 GCGCCCTCTCTGGGGCACCCACAGCGACCGCCCTCGGTGAGCTCCACAAAGGAGTCTCG 6261
Qy 1428 -IleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLys 1447
Db 6260 GCTCCCGCGTACCGATGG-----CTGCACAGCGCCGAGCGAGACACGCCCATTTCCG 6204
Qy 1447 sGluGlySerIleThrGlnGlyThrProLeu-LysTyrAspThrGlyAlaSerThrThrG 1467
Db 6203 ATGGAGGGGTGTGAGGATGCGCTCTGCGACAGCCGAG----- 6163
Qy 1467 lysrLysLysHisAspValArg-----SerLeuI 1477
Db 6162 -----AAGACACCGCGCCCGCAGCGGCACCGCCCGCTCGTGGTGGCGAGCAGT 6108
Qy 1477 leGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAla 1497
Db 6107 TCGCGATCCGGGCGCGCGGAGCGGTTCACGGCCGAGCAACTCGACCTCGGCGCCC 6048
Qy 1497 rgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGly----- 1513
Db 6047 CTGCGAATGAGCGTCTGT-----GATGCTGCGCAAGCTCCGAGCGCGGTCCCGCGG 5991
Qy 1513 ----- 1513
Db 5990 TCGCGCATGATAGCCAGCGGTGGTGAGGGGTGCGGGGTCTGCTCGGCAACGGTTTC 5931
Qy 1514 -----ThrAlaSerSerSerGlyGlySerIleAlaArgGly- 1525
Db 5930 CAGGTGATCCGATACCGCAGCCACCGCGGTGCGGGTGGTGGTGTGCGCGTGGCG 5871
Qy 1526 -----AlaProValIleValPro----- 1531
Db 5870 GGGTTGTGCGGGTGGTGGAGTCGGTCCCGAGTAGTGGCGGTTCGAAACGGATAC 5811
Qy 1532 -----GluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProp 1550
Db 5810 GTCGGCAGGTCAACCAACCGCGCGCCGCC-----CGCGAACACACACCTTCCAG 5760
Qy 1550 heAlaGlyHisLeuProArgGlySerProValThrMetArg-----GluProThrPro- 1567
Db 5759 TCAACTCAT-----CCGCGCAGCAACGCTCTACCCACCGAGTCAAAAACCGGCCAC 5703
Qy 1567 ----- 1567
Db 5702 CCACCATCATCAGGGCGAGGACCCACCCACCGCGCGCATCCGGATCCACACC 5643
Qy 1568 -----ArgLeuGlnGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuT 1586
Db 5642 GCCACGACTCCAAAC-----CCACCGGCCAAACCGGATGCCCGGAACACTCC 5592

Qy 1586 hrSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHisHisPro- 1605
Db 5591 ACAAAACCCGAAACCCATCCCAACAAGCGCCACCACATCCGAAAAACAACACCCGC 5532
Qy 1606 -----HisProIleSerProTyrGluH 1613
Db 5531 TCACGCAGATTCCGATACCAATACCACCATCCAAACCCACCCCATCAACACACACCC 5472
Qy 1613 isLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheA 1633
Db 5471 ACCACCGACGA-----ATAAACCGGCACACACCCCGCCCGGACAC 5430
Qy 1633 spProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProA 1653
Db 5429 AAACACCCAAACACCCCGACGAGCATCCCTCCACCGCTCCACATCGCGCATGGAC 5370
Qy 1653 rgHisLeuAlaProAsnProThrTyrProHisLeuTyrPro----- 1666
Db 5369 GCATAATCCACCGGCACCGCGCACACACACCCATCCCGCTCACACTACCAACAAC 5310
Qy 1667 -----ProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgg 1683
Db 5309 CCCACAGCAGCATCCACATCCCGGCACACACCGCTCAACGACGAGCATTCACCGCGCC 5250
Qy 1683 lnThr-----llelleAsnAspTyrIleThrSerGlnG 1694
Db 5249 ACACCCAAACCGACCCCGCCCAACGGTCTCTCGCATCCAGTCGGCAGTCCG 5190
Qy 1694 lnMetHisHisAsnThrAlaThr-AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeu 1713
Db 5189 ACCGACACCATTCGCACACCGCGCAACACCTTACCAATACCGCGACCGCAACGCC 5130
Qy 1714 SerProArgGluSerSerLeuAlaLeuAsn-TyrAlaAlaGlyProArgGlylleleAs 1733
Db 5129 ACCACCGCGCGCCATCAGCCACACTCAACACACCCCGCCACCG----- 5084
Qy 1733 pLeuSerGlnValProHisLeuProValLeuValProProThr-----Pr 1748
Db 5083 -----CGCAGCAACTCAG-----CCTGGAAATGACCCACCGCATCGGACACACCC 5031
Qy 1748 oGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnPro-- 1767
Db 5030 ACCGACATCCACAACCGCG-----CCACGACACCATCACCGCC 4992
Qy 1768 -----PheSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHisLe 1785
Db 4991 CACGACACCGGCTGCACCATCAACCGCACCC-----ACACCCCTCACCAGCGCCACCC 4935
Qy 1785 uThrLysProThrThrThrSerSerSerGluArgGluArgAspArgGluArgAs 1805
Db 4934 CGCACACCTCCAAATAATCCCAATCCCA----- 4904
Qy 1805 pArgAspArgGluArgGluLysSerIleLeuThrSerThrThrValGluHisAlaPr 1825
Db 4903 -----ACCCGACAAACCGCGCACCATCTCC 4878
Qy 1825 oIleTrpArgProGlyThrGluGlnSerSerGlySerSerGlySerGlyGlyG 1845
Db 4877 A-----CCATCGACTCAGAAAC-ACCTCGACTCTCCAGCAGCGCGCCACCAT 4828
Qy 1845 yGlySerSerArgProAlaSerHisSerHisAlaHisGlnHis-SerProIleSerP 1865
Db 4827 ACCACCCACTGCCAACCTTG-ACCCGAAACAAACAC-----CACCCGACACCCAC 4772
Qy 1865 roArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysG 1885
Db 4771 CAGCAACACCGGACCGCACCGAACCC----- 4746
Qy 1885 lyIleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSerS 1905
Db 4745 -----ACCACCGACCGAACCGGAA-----GCCACCGCGCGGAAACACCGCATCCA 4700

1905 exProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrL 1925
1925 eu-AspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArg 1944
4699 AACCGCCACCATCTGGGAGACCCCAACACG-----ACCG 4658
1925 eu-AspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArg 1944
4657 CCGGATGCTGAACACCGACCGGACGACACCAACGACCAACCAATTCACCGACCGG 4598
1945 ValAlaArg-----ProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLys 1962
4597 CACCATCGGAAGACCGCGCACCTCAGCA---ACCGCGCGCATAGCGCGCAAG 4541
1963 ProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArg 1982
4540 CCGCGCTCGAAC-----GACCGGAACACCAACCGGACCAACCGATAGACCCCTAC 4490
1983 Pro-----LeuValProPro 1987
4489 CCACACGACGATCGGTGCTCGGTAAACACCGGATCGTGAATTCGGGGGCGCTGCT 4430
1988 ValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAla 2007
4429 CCAAAATCAGATCGGATTCGTCCTCCGACACCCCAACGCGGACACACCGCGACGCA 4370
2008 SerProAspProAlaProAla-----SerAlaSerAspProHisArgGluLys 2025
4369 CACGACCCACCGCGCAATCCGCGCTCCGTCACCAACTCCACCCAC---CCGAAG 4313
2026 ThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGly 2045
4312 ACCAATCCACCAACCGCGACGCA-----CATCCA 4283
2046 SerSerTyrSerProGluGlyValGluProValSerProValSerSerProSerLeuThr 2065
4282 CATGCAACGTGCGCGGACACACCCCGCGCGACGCA-----TCACCATCTGATCA 4229
2066 HisAspLysGlyLeuProLysHisLeuGlu-LeuLeu----- 2077
4228 CAC-----CGGCACACCGCGACGCGCTGCGTATGACCAATATTCGACTTCA 4181
2078 -----AspLysSerHisLeuGluGlyGluLeuArgProLysGln----- 2090
4180 CCGAACCCAAACACCGCCCAACCGACGACCCCGCCCATACGACCCCAACACGCGCT 4121
2091 -----ProGlyProValLysLeuGlyGlyGluAlaL 2101
4120 CAGCCTCAATCGGATCACCAACGAGTCCCGTCCATCGGCTC-----CA 4073
2101 aHisLeuProHisLeuArg-----ProLeuProGluSerGlnProse 2115
4072 CCACAT-CCACATCCGACACACCAACCGCGCGCCACCAACGCGCAACCAATCACCGCA 4014
2115 rSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValThrLe 2135
4013 CGCTGCGCCA---CCCCGACGCGCGCG----- 3989
2135 uAlaGlnHisLeuSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLe 2155
3988 -----TCMAACCATTCGACCCACCATCTCTGATTCACC 3957
2155 userAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAs 2175
3956 GCAGAACCCCGCACACCGGCAACA-----CCCGATGACCCACGCTCCCGG----- 3911
2175 pLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaAr 2195
3910 -----CCACCGAACACCGCTCCCAACACACCAACCCCGCACCTCCCGCCACCCC 3861
2195 gGlySerProHis-----SerGluGlyG 2203
3860 GTCCCATCCGACACACCGGCAACCGCGGACACCGCCCATCATCCCGCAACCCACCTGCG 3801
2203 yLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAspGlyIleG 2223

3800 CGAGAAATCTCACAAACCAACCCCGCG----- 3773
2223 uProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrPr 2243
3772 -CCGNAATCACCGTCACCCACCGCAACGCAATCATCACTCACCACCGCAACCC 3714
2243 oLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProG 2263
3713 T-----GACACGCCAAATGCAACGCCCAACGACGACGACCAACCGCGTATCAAC 3663
2263 yAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetVa 2283
3662 GAAATCGACGCGACCTCCCAAC----- 3641
2283 lLysSerLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluPro-- 2302
3640 -----CCAACACATACGAAACACGACCCGAC 3615
2303 -GluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyTh 2322
3614 ACCACACTCGCGGACGTCCTCCGTCAACA-----ATAACCT----- 3578
2322 rGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLe 2342
3577 -----CAAACCCACACCCCGACT----- 3560
2342 uGluAlaIlelleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluSerProPr 2362
3559 -----CCACC----- 3555
2362 oLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetPr 2382
3554 AACCGGCGCCATACCTCTGACCCA---TCACCCCAACACACCGCGTCCGCAACCC 3498
2382 oIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyL 2402
3497 CGCAAAACCCCAACCGGAT---CAATCCCGCGACGCTCCACGCGCTCC----- 3455
2402 sAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAl 2422
3454 -----ACGACACTCCAAAAACAACGCTGCTGGGATCCATCGCCACC-GCCTC 3406
2422 aSerGlyAspArgPro----- 2427
3405 ACGCGCGACACACCAAAAAAACTCCGCAATCCCGCGCTCATACACAAACCCCCC 3346
2428 -----ProSerValse 2431
3345 CTCCTCGCATACGACTTCCCCACCAACCGCGCATCAACAAACCCCTCCACATC 3286
2431 rSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpG 2451
3285 CCAACCCCGATCCACCGGAACCCCGACACCGCATCCCAACCCCAAC---ACCAATC 3229
2451 uAspArgProSerSerAlaGlySerThrProPheProTyr-AsnProLeuIleMetArgL 2471
3228 CCACAAACCTCCGCAACACCAACCCCAACCGGAAAAACGAAACCA----- 3182
2471 euGlnAlaGlyValMetAlaSerProProPro-----G 2483
3181 -----TCCCCACCAACCAACCGGATCATCCACGACGACGCGGA 3142
2483 yLeuProAlaGlySerGlyProLeuAlaGlyProHis----- 2495
3141 GGTGACTCCGCGACGTTCTCCGTTGCCGCGCTGAAATTTCTCTGACGCTTCCGGC 3082
2496 -----HisAlaTrpAsp 2499
3081 AGCGGCAATCGGGTGGAT 3062
RESULT 69
AX697977/c

QY	573	AsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaIleSer	592
Db	14778	ACCGAAGCCGACACACACCGGAAACCGACGAGCTCTGCAACACACACCGCTCG	14719
QY	593	GluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSer	612
Db	14718	CCGCCATCCACGACACCGGCATCGGAATCGGAA-----	14686
QY	613	SerArgTrpThrGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArg	632
Db	14685	-----TCGGAATCGGAATCGGAATCGGCA-----	14662
QY	633	AsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsn----	651
Db	14661	-----GGATCGTCGAATCTCGGGGCGCTCTCCAAATC	14629
QY	652	-----PheTyrPheAsnTyrLysLeuArgGluAsnLeuAspGluIleLeuGln	668
Db	14628	AGATGCGCATTCGTGCC-CGACACCCCAACCGCGACACACCCGCGACGACACGACC	14570
QY	669	HisLysLeuLysMetGluLysGluArgAsn-----	678
Db	14569	CACCACCGGCATCCCGCGCTCCGTACCAACTCCACCCACCGCAACCAATCCAC	14510
QY	679	-----AlaArgArgLysLysLysAlaProAlaAlaAla	690
Db	14509	CAACGGCGAGCGCATCCATCATCAACGTCGCGGCA-CACCCCGCGCGCAACGCCA	14451
QY	691	SerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGlyVal	710
Db	14450	TCACCATCTTGATCACACCGCGCACACCGGACCGCGCTCGTATGACCAATTCGACT	14391
QY	711	SerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGlu	730
Db	14390	TCACCGAACCACACACACCGCCCCACCCACGACCGCGCCATACGACC-----	14340
QY	731	ValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAsp-----	748
Db	14339	-----CCAAACACCGCTCAGCTCAATCGATCAATCGATCAATCGATCAATCGAT	14304
QY	749	GluSer-----IleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsn	765
Db	14303	GAGTCCCGCTCCATGCGCTCCACCATCCACAT-----CCGACACACCCCAAC	14253
QY	766	GlyProLysProProAlaThrLeuGlyAlaAspGlyProProGlyProProThrPro	785
Db	14252	CGCGCGCACCAACCGCGCACCAATCACCGACGCTGCGCA-----CCCGGACGCG	14199
QY	786	ProArgArgThrSerArg-----AlaProIleGluPro	796
Db	14198	CCGTCAAAACCATTCACCCCATCTGATTCACCGGAGAACCCCGGACCA-----CCGCA	14142
QY	797	ThrProAlaSerGluAlaThrGlyAlaProThrProProAlaProProSerProSer	816
Db	14141	ACACCCGATGACCGCTCCCGCGCCACCGACACCGCTCCCAACACACACACCCCGCAC	14082
QY	817	AlaProPro-----ProValValProLysGluGluLysGluGluThrAlaAlaAlaPro	835
Db	14081	CCTCCCGCCACCGCTCCCATCCGACACCCCGCAACCGCGCACCGCCCATCATCCCG	14022
QY	836	ProValGluGluGlyGluGlu-----GlnLysProProAlaAlaGluGlu-----	850
Db	14021	ACAACCCACCTGCGCGAGAAACTCCAAACACCCCGCGCGCGCATCACCGTACCCC	13962
QY	851	-----LeuAlaValAspThrGlyLysAlaGlu	859
Db	13961	CACCCGCCAACCCACATCACACTCACCAACCGCAACCCCTGACACCGCAAAATGCAACG	13902
QY	860	GluProValLysSerGluCysThrGluGlu-----	869
Db	13901	CCACCAACGACGACCAACGCGCTATCAACCGAATCGACGACCCCTCCCAACCCCAACA	13842
QY	870	--AlaGluGluGlyProAlaLysGlyLysAsp-----AlaGluAlaAlaGlu	884
Db	13841	CATAGAAACAGCCCGACCGAGCTCACCGACTTCCGGTCCGAGAGATAGCCCTCGA	13782
QY	885	AlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThr	904
Db	13781	GTTCCTCGGAGCGCTCGTCAGACGCGCGCATATCTCTGCTGGTTCAGGCGCGAGTACA	13722
QY	905	ThrAlaLysSerSerGlyAlaProGluAspSerAspSerSerAlaThrCysSerAlaAsp	924
Db	13721	-----CCCGCTCGGCTACCCGACAC-----CCGCGCGAT	13689
QY	925	GluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeu	944
Db	13688	CGATTCGCCGACGCT-----CCAACGCTCCACG	13659
QY	945	LeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLys	964
Db	13658	ACGTCTCTCAACAAACGAGCGCTCTCGGATCCATCGCACCGCTCACGCGCGACACAC	13599
QY	965	Gln-----LeuLysGlnArgAlaAlaAlaIleProProIleGlnVal	978
Db	13598	CAAGACTCCGCATCGAACATCCCGGCTCATACACAAACCCCTCCGCGCATAGC	13539
QY	979	ThrLysValHisGluProProArgGluAspAlaAlaProThrLysProAlaPro-----	996
Db	13538	ACTTCCCGCGCGCGCGACCGGATCAAGAGAGAGTCCACGTCGCCCGCGCTCGT	13479
QY	997	-----ProAlaProProProProGlnAsn-----	1004
Db	13478	CGGGGCTGGGCGCATGACGTACAGCACCGCGCACGAGGAGTCCAGAGCTTCTCGGAG	13419
QY	1005	-----LeuGlnProGluSerAspAlaPro-----	1012
Db	13418	TGTGCGCTCCGCGGGTAGCGCGACGCTCCCGACGATGGCAATAGCTCTCGACGC	13359
QY	1013	-----GlnGlnProGlySerSerProArgLysSerArgSerProAlaPro	1028
Db	13358	GCACGCTCGCGCAGTCACCGGA-----CTTCTCGGACAGAGCGCGCGACGAGCGCT	13302
QY	1029	ProAlaAspLysGluAlaPheAla-----AlaGluAlaGln	1040
Db	13301	CCAGGAATTCGCAACAGTGGACCGGGTGGGTGGTTCGAGAACCCCTCGGAGGCGAC	13242
QY	1041	Lys-----	1041
Db	13241	TCAGACCGCTCGTCTGAGTGTTCGCGAGTTCGAGCGCGCTGAGCGAGTTCGAAGC	13182
QY	1042	-----LeuProGly-----AspProProCysTrpThrSerGlyLeu	1053
Db	13181	CGCGCTCTTGAAGCGCGGGTGGGCGCGACCGCATCCACCGCTCTGCTCGCTCAGGACG	13122
QY	1054	ProPheProValProProArgGluValIleLysAlaSerPro-HisAlaProAspProse	1073
Db	13121	CCGCGCGCGCGCGC-----GCACCGTTTCGAGGAGGAGGACCGCT	13083
QY	1073	rAlaPheSerTyrAlaProProGlyHisProLeu-----ProLeuGlu	1087
Db	13082	GCGTTCGCGCTCGGCGCATCCCGCGCACCGCTCGCGCAGCGCGCGCTCGCGACTCGG	13023
QY	1087	YLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnPro-----	1105
Db	13022	TGTCCAC-----TTTCCGGCATCGAGCGCGCTCTGCGCTCGGGCACCCCGCTCGA	12972
QY	1106	-----ProProLeuIleSerSerAlaLys-----	1113
Db	12971	AGACCCCGCGCGAGCTCGCGAGAGCGGAGCGGAGGCGGAAACCGTCCAGTCAGCTCGG	12912
QY	1114	-----HisProSerValLeuGlu-----	1119
Db	12911	CGACGCTGACAGACACGCTCATCGCGTCGAGGCGCTCTGCGAGCGCGCGATGCCGAGCT	12852
QY	1120	-ArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGlu	1139

QY 1694 lMethHisHisAsnThrAlaThr-AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeu 1713
Db 10890 ACCGACACCATCGCACACACCGCCGACACCATCTACCAATTAACCGCGCGACCAACGCC 10831
QY 1714 SerProArgGluSerLeuAlaLeuAsn-TyrAlaAlaGlyProArgGlyLeu 1733
Db 10830 ACCACCGCGCGCCATCAGCCACATCAACACACCCGCCACACCG 10785
QY 1733 pLeuSerGlnValProHisLeuProValLeuValProProThr-----Pr 1748
Db 10784 ----CCGACCAACCTCAC---CTCGAATGATGACCCACACCGCATCTCCGCGCACACACCC 10732
QY 1748 oGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnPro--- 1767
Db 10731 ACCGACATCCACACCGCG-----CCAACGACACCATCACCGCC 10693
QY 1768 -----PheSerSerArgHisSerSerProLeuSerProGlyGlyProThrHisLe 1785
Db 10692 CAGACACCGCGCTGCACCATCAACCGGACCC---ACACCCCTCACCCTCCGCGCCACCC 10636
QY 1785 uThrLysProThrThrSerSerSerGluArgGluArgAspArgGluArgAs 1805
Db 10635 CGCACCACTCCAAACAAATCCCAATCCACAA----- 10605
QY 1805 pArgAspArgGluArgGluLysSerLeuThrSerThrThrValGluHisAlaPr 1825
Db 10604 -----ACCCGACACACCGCCGACACACTCC 10579
QY 1825 oIleTyrArgProGlyThrGluGlnSerSerGlySerSerGlyGlyGlyGly 1845
Db 10578 A-----CCATCGACTCAGAAAC-ACCTCGACTCTCCAGCAGCGCCGCCAT 10529
QY 1845 yGlySerSerSerArgProAlaSerHisHisAlaHisGlnHis-SerProLysSerP 1865
Db 10528 ACCCACCACTGCGCAACCTG-ACCCGAAACACAAACAC---CACCCGACCAACCCAC 10473
QY 1865 roArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysG 1885
Db 10472 CAGCAACACCGCGCCACCGAACCC----- 10447
QY 1885 lylleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSerS 1905
Db 10446 -----ACCACACCGAACCCGAA-----GCCACCGCCCGGAAACACCGCATCCA 10401
QY 1905 erProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrL 1925
Db 10400 AACCGCGCACCATCTCGCGCACACCCCGCAACACG-----ACCG 10359
QY 1925 eu-AspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArg 1944
Db 10358 CCGCATGCTGAAACACCGCGCGAGCAGACCAACGACCAACCCACATTCACGACCGG 10299
QY 1945 ValAlaArg-----ProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLys 1962
Db 10298 CACCATCGGAAGACCGCGCACACTCACGCA---ACCGCGCGCATAGCGCGCAAG 10242
QY 1963 ProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArg 1982
Db 10241 CCGCGCTCGAAC-----GACCCGAAACCAACCCAGCGGAACCAACCGATAGACCCCTAC 10191
QY 1983 Pro-----LeuValProPro 1987
Db 10190 CCACACGACATCGGTGCGGTGCGTAACACCGGATGCTCGAATTCGGGGCCCTGCT 10131
QY 1988 ValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAla 2007
Db 10130 CCAAAATCAGATGCGCATCTGCTCCCGACACCCCAAGCGCGACACACCGCGCACGCA 10071
QY 2008 SerProAspProProAlaProProAla-----SerAlaSerAspProHisArgGluLys 2025
Db 10070 CACGACCCACCAACCGCGCAATCCCGCGCTCCGTCAACCACTCCACCCAC---CCGAAG 10014

QY 2026 ThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGly 2045
Db 10013 ACAATCCACCAACCGCGAGGCA-----CATCCA 9984
QY 2046 SerSerTyrSerProGluGlyValGluProValSerProValSerProSerLeuThr 2065
Db 9983 CATGCAACGTCCGCGCACACCCCGCGCGCAACGCCA-----TCACCATCTTGATCA 9930
QY 2066 HisAspLysGlyLeuProLysHisLeuGlu-LeuLeu----- 2077
Db 9929 CAC-----CCGCCACACCGCGCGCTCGGTATGACCAATATTCGATTCA 9882
QY 2078 -----AspLysSerHisLeuGluGlyGluLeuArgProLysGln----- 2090
Db 9881 CGAACCCCAACAAAGCGCCACCCACACGAGCCCGCGCCATACGACCAATATTCGATTCA 9822
QY 2091 -----ProGlyProValLysLeuGlyGlyGluAlaAl 2101
Db 9821 CAGCCTCAATCGGATCACCAACCGAGTCCCGTCCATCGCGCTC-----CA 9774
QY 2101 aHisLeuProHisLeuArg-----ProLeuProGluSerGlnProSe 2115
Db 9773 CCACAT-CCACATCCGACACACCAACCGCGCGCCACCAACCGCGCCATCACCCGA 9715
QY 2115 rSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValThrLe 2135
Db 9714 CGCTGCGCA---CCCCGACGCGCGCG----- 9690
QY 2135 uAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLe 2155
Db 9689 -----TCAAAACCATTCGACCCACCATCTCTGATTCA 9658
QY 2155 uSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAs 2175
Db 9657 GCAGAACCCCGCGACACCGCGCAACA-----CCGATGACACCGCTCCCGCG----- 9612
QY 2175 pLeuArgArgProProSerAspLeuTyrLeuProProAspHisGlyAlaProAlaAr 2195
Db 9611 -----CCACCGACACCGCTCCAAACACCAACCAACCGCGCACCTCCCGCCACCC 9562
QY 2195 gGlySerProHis-----SerGluGlyG 2203
Db 9561 GTCCCATCCGCGACACCGCGCAACCGCGCCATCACCCGACCAACCCACCGCTGC 9502
QY 2203 yLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleG 2223
Db 9501 CGAGAAATCTCCAAACAAACACCGCGCG----- 9474
QY 2223 uProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrPr 2243
Db 9473 -CCGCAATCACCGTCAACCGCGCGCGCGCACATCACATCACCAACCGCAACCC 9415
QY 2243 oLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProG 2263
Db 9414 T-----GACACGCGCAATCAACCGCGCGCGCACACGACGACCGCGGTATCAAC 9364
QY 2263 yAsnThrSerGlnProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetVa 2283
Db 9363 GAAATCGACGCGCGCTCCAAAC----- 9342
QY 2283 lLysSerLysLeuGlnIleAsnLysLysLeuAsnThrHisAsnArgAsnGluPro--- 2302
Db 9341 -----CCAAACATACGAAACACGACCGCGAC 9316
QY 2303 -GluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyTh 2322
Db 9315 ACCACACTCGCGCGCGTCCCGCGTCAACA-----AATAACCT----- 9279
QY 2322 rGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLe 2342
Db 9278 -----CAACCCACCGCGACT----- 9261
QY 2342 uGluAlaIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProPr 2362

```
Db 9260 -----
Qy 2362 oLeuSerAlaAenAlaPheAenProLeuAenAlaSerAlaSerLeuProAlaAaMetPr 2382
Db 9255 AACCGGGCCCTACTCTCTGACCA---TCACCCCCACATACACCGCTCCGCGAACCC 9199
Qy 2382 oileThrAlaAlaAaspGlyArgSerAaspHisThrLeuThrSerProGlyGlyGlyLy 2402
Db 9198 CGCAAAACCAACGGAT---CAATCCCGCAGCTCCCAAGCGCTCC----- 9156
Qy 2402 sAlaLySvalSerGlyArgProSerArgLyAlaLySserProAlaProGlyLeuAl 2422
Db 9155 -----ACGACACCTCCAAAACAAACCGCTGCTCGGATCCATCGCCACC-GCCTC 9107
Qy 2422 aserGlyAaspArgPro----- 2427
Db 9106 ACGCGGGGACACACCAAAAAAATCCGGATCGAATCCTCCCGGTCATACACAAACCCCCC 9047
Qy 2428 -----ProSerValse 2431
Db 9046 CTCCCGCACATACGACTTCCCCACACACCCCGCAGCGATCAAAACACCGCTCCACATC 8987
Qy 2431 rSerValHisSerGlyAaspCysAenArgArgThrProLeuThrAenArgValTrpGl 2451
Db 8986 CCAACCCCGATCCACCGGAACCCCGACACCGCATCCCGCCCAAC---ACCAAAATC 8930
Qy 2451 uAaspArgProSerAlaGlySerThrProPheProTyr-AenProLeuIleMetArgL 2471
Db 8929 CCACAAACCTTCCCGACACACACCCACCCCGGAAACAGCACACCCA----- 8883
Qy 2471 euGlnAlaGlyValMetAlaSerProProPro-----G 2483
Db 8882 -----TCCCCACACACACCGATCATCCACGAGCAGCGCGA 8843
Qy 2483 lyLeuProAlaGlySerGlyProLeuAlaGlyProHis----- 2495
Db 8842 GGTGACCTGCGGACGTTCTCCGGTTCGCGGCCCTGAATTTCTCTCGACGTTCCCGGC 8783
Qy 2496 -----HisAlaTrpAasp 2499
Db 8782 AGCGGCCATCGGGTGGAT 8763

RESULT 70
LOCUS AC134881 204006 bp DNA linear HTG 01-OCT-2002
DEFINITION Homo sapiens chromosome UNK clone RP11-732A22, WORKING DRAFT
SEQUENCE, 18 unordered pieces.
ACCESSION AC134881
VERSION AC134881.1 GI:23396432
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
WATERSTON, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 204006)
AUTHORS WATERSTON, R.H.
TITLE Direct Submission
JOURNAL Submitted (01-OCT-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: H_NH0732A22
```

```
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-terminator ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 198120 bases at least Q40
Consensus quality: 199344 bases at least Q30
Consensus quality: 200029 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1393: contig of 1393 bp in length
* 1393: gap of unknown length
* 1493: contig of 1257 bp in length
* 2751: gap of unknown length
* 2851: contig of 1617 bp in length
* 4468: gap of unknown length
* 4567: contig of 1370 bp in length
* 5937: gap of unknown length
* 6038: contig of 1794 bp in length
* 7931: gap of unknown length
* 7932: contig of 2466 bp in length
* 10398: gap of unknown length
* 13591: contig of 3093 bp in length
* 13591: gap of unknown length
* 15421: contig of 1731 bp in length
* 15422: gap of unknown length
* 19656: contig of 4135 bp in length
* 19756: gap of unknown length
* 19757: contig of 6304 bp in length
* 26061: gap of unknown length
* 26161: contig of 5527 bp in length
* 31687: gap of unknown length
* 31688: contig of 4306 bp in length
* 36093: gap of unknown length
* 36193: contig of 9974 bp in length
* 46167: gap of unknown length
* 46168: contig of 12141 bp in length
* 58408: gap of unknown length
* 58509: contig of 71246 bp in length
* 129754: gap of unknown length
* 129755: contig of 72829 bp in length
* 129854: gap of unknown length
* 202683: contig of 72829 bp in length
* 202684: gap of unknown length
* 202784: contig of 1044 bp in length
* 203827: gap of unknown length
* 203828: contig of 79 bp in length.
* 203928: Location/Qualifiers
1. .204006
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="UNK"
/clone="RP11-732A22"
1. .1393
/notes="assembly_name:Contig116"
1494. .2750
/notes="assembly_name:Contig119"
2851. .4467
/notes="assembly_name:Contig120"
4568. .5937
/notes="assembly_name:Contig122"
6038. .7931
/notes="assembly_name:Contig123"
7932. .10397
/notes="assembly_name:Contig124"
```

[illegible]

```
Qy 547 LysGluLysThrAspAspThrSerGlyGluAsnAspGluLysGluAlaValAlaSer 566
Db 146500 -----GCGAGC 146505
Qy 567 LysGlyArgLysThrAlaAsnSerGlnGlyArgGlyGlyArgLleThrArgSerMet 586
Db 146506 AAGCTGGCCAAACAGCCAAAGCCAGTCAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAA 146565
Qy 587 AlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSer 606
Db 146566 GCCAGACAAGCC---AGCCAAACAGCCAGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC 146622
Qy 607 MetGluLeuAsnGluSerSerArg---TrpThrGluGluGluMetGluThrAlaLysLys 625
Db 146623 CAAGAGACCCCAAGCAGCCAGCAAGCTGGCCAAACCCAGCCAGCCAGCCAGCCAGCCAA 146682
Qy 626 GlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThr 645
Db 146683 GCC----- 146685
Qy 646 ValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIle 665
Db 146685 ----- 146685
Qy 666 LeuGlnGlnHisLysLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLys 685
Db 146686 -----AAGCCGGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC- 146732
Qy 686 AlaProAlaAlaSerGluGluAlaAlaPheProValValGluAspGluLumet 705
Db 146733 AGCCCTGCCAGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC 146792
Qy 706 GluAlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHis 725
Db 146793 CAAGCCAGCCA-GCCAGCCAGCAGC-----CAAGCCAGTCAGCCAGCCAGCCAA 146839
Qy 726 AlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSer 745
Db 146840 GCCAGC-----GAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAA 146878
Qy 746 SerAspThrGluSerIlePro-----SerProHisThrGluAla 758
Db 146879 GCCAGCAAGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 146938
Qy 759 AlaLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGly--- 777
Db 146939 AGCCAGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC 146998
Qy 778 -----ProProProGlyProProThrProProArgArgThrSerArgAlaProile 794
Db 146999 CAGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC 147058
Qy 795 GluProThrProAlaSerGluAlaThrGlyAlaProThrProProProAlaProProSer 814
Db 147059 CAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC 147115
Qy 815 ProSerAlaProProValValProLysGluLysGluGluThrAlaAlaAla 834
Db 147116 CAGTCAGCCAGCCAGC---CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 147172
Qy 835 ProProValGluGluGlyGluGluGlnLysProProAlaAlaGluGluLeu----- 851
Db 147173 -----AAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC 147214
Qy 852 ---AlaValAspThrGlyLysAlaGluGluProValLys---SerGluCysThrGluGlu 870
Db 147215 CCAGCAACCCCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 147274
Qy 870 laGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGly 890
Db 147275 CGAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC 147334
Qy 890 laLeuLysAlaGluLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerG 910
```


Qy	1445	ProLeuLeuGluGlySerIleThrGlnGlyThrProLeuLeuTyrAspThrGlyAlaSer	1464
Db	2636	CGG-----AGCTCTACCCCGCCACTACCGCGATCATTCAGGCTCTTCTCTCTCC	2686
Qy	1465	ThrThrGlySerLysLysHisAspValArgSerLeuIleGlySerProGlyArgThrPhe	1484
Db	2687	AGCGGCCCTCGTCG-----	2701
Qy	1485	ProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyr	1504
Db	2702	CTCTCTCCGTCGCCCT-----	2716
Qy	1505	GluGluSerLeuLysSerArgProGlyThrAlaSerSerSerGlySerIleAlaArg	1524
Db	2717	-----ACCCGCGCCTCGACTCGCTTCTGGTCT-----	2743
Qy	1525	GlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGlu	1544
Db	2744	-----CCTGCTCCGTCGCCAGCGGAATCCAAAGAGACCTCT-----	2782
Qy	1545	AspHisGlyAlaProPheAlaGlyHisLeuPro-----ArgGly	1557
Db	2783	-----GGGCAACCCCTTGGCAATCTTCCCCCGAGTATCTCGTCTCCGACCTTGCC	2836
Qy	1558	SerProValThrMetArgGluProThrProArgLeuGlnGluGly-----SerLeuSerSer	1576
Db	2837	TCTCCACCAACCATCTCACCGCCCTCGCTTCACTTACCAGTGGTGTCACTCTCCGCT	2896
Qy	1577	SerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerPro	1596
Db	2897	CAGGAGAGTGGTCCACGCGCACCGGCTCTCTCCGCGCCCTTTCGCTTTGAGATCACCC	2956
Qy	1597	HisSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArg	1616
Db	2957	-----ACACCCCAACCGTCTCTCCGCGCACCG-----TCG	2986
Qy	1617	GlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSer	1636
Db	2987	GGCACTTCGCGCCTCTCGCGCCCGCCTCACCCCGCACTTCGCTGCTCTCACCGGGGGCGCC	3046
Qy	1637	IleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProArg-----HisLeu	1655
Db	3047	CTTCCCAAGTCACTCCGGC-CCCAAGCCTCCGCG-TCCTCCCGCGCGCGCTTT	3102
Qy	1656	AlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyr-----	1673
Db	3103	GGGCTTCGCGCTGAAGCTCCCGCG-CTCTCCCTGACCCCGCTGGCTTCACTCCC	3156
Qy	1674	-----ProAspThr-----AlaAlaLeuGlu-----	1680
Db	3157	ATCACCGAGCCACCCCTGCCACATCCCGCTTCGCGCTCGCGCGCTTCGCTTCGGCTC	3216
Qy	1681	AsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMethHisAsnThrAla	1700
Db	3217	AACCGCCAGCC-----CTCCCGATCCGCAAGTCCAC-----GCC	3252
Qy	1701	ThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeu	1720
Db	3253	TACTCCATCGCC-----CTCCCTCGCGCAAAAACCTC	3285
Qy	1721	AlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeu-SerGlnValPro-----	1738
Db	3286	GTCTCCATCTCAAGAACGGATTGACGGTCTCGTCTCTCTCTCCCGCTCCGACCGC	3345
Qy	1739	-----HisLeuProValLeuValProProThr-----	1747
Db	3346	ACCAACCTCTCTCCCTCATCCAGCCCTTGATCACCGCGGACTTCCCTCCGCGCCGC	3405
Qy	1748	----ProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGln	1766
Db	3406	CCCCCGCTCGGCGCTCATCCACATCGCGCGCTTGGCTGCGGTAAAGAGTACCCCATC	3465

Qy	1766	nProPheSerSerA	gHisSerSerProLeuSerProGlyGlyProThrHisLeu	--	1785
Db	3466	CAGCAACTGCTGCCACCCAGACCTTC	CGCGCGTGCACCCACCACCTGAG	3525	
Qy	1786	-----ThrLysProThrThrSerSerGlu	-----	1795	
Db	3526	CTTCGGCCAGATGNAAGCTCGCTCAAGCT	CGAAGGCCCTCCTCTTGGCGGTCTCC	3585	
Qy	1796	-----ArgGluArgAspArgAspArgAspArgAspArgGluArgGluLysSe	1813		
Db	3586	ACCTGGGAGACAGCCCTCGCAAGCGTGT	CGTCTGCGTCTTGTATCGACGAGGTCTACAAG	3645	
Qy	1813	rLeuLeuThrSerThrThrValCuiHisAlaProIleIleTpaArgProGlyThrGluGl	1833		
Db	3646	CTCCCTCGCGGGTACCTCGACCTTGGCCCT	TCGCGC-----ACCCACC	3690	
Qy	1833	nSerSerGlySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerA	gProAlase	1853	
Db	3691	GTCCAGTTCGTATCATCTCTCGGACCCCT	CCCAAGGATCTACAAGCCACCA	3745	
Qy	1853	rHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnGl	1873		
Db	3746	-----ACCCCGACTCTTCCACACCGGCTCAT	CCCCCGAAGA	3783	
Qy	1873	nArgProSerValLeuHisAsnThrGlyMetLysGlyLleIleThrAlaValGluProSe	1893		
Db	3784	GACCACTTCGCGGTTCATGACTTCTAT	GTCTGTGGACTCGCGCTCCCGCGCTC	3843	
Qy	1893	rLys-----ProThrValLeuArgSerThrSerThrSerSe	1905		
Db	3844	GTCCGCGACTTCTTCGGCGTACCACCA	CCCCCACTCGCGGTACCTCGCTCGCTCGCC	3903	
Qy	1905	rProValArgPro-----	-----	1909	
Db	3904	TCCTCTCAACACCACTCAATCCCTCTCTCTCGTCCCTCCGACTCCGACTGSCCGAGCCCTC	3963		
Qy	1910	-----AlaAlaThrPhePro-----ProAlaThrHisCy	1919		
Db	3964	ACTGTCGGCGGCCACCGCGCATCACCTACGCGCCTCCCAAGTTTCCA	ATTACCCCGCC	4023	
Qy	1919	sProLeuGlyGlyThrLeuAspGlyValTyr-----ProThrLeuMetGluProVa	1936		
Db	4024	CCCGTCCATCTTCTCTTGACCGCAATTTAACTCGTCACCAACCATGTCGCCCTCGTC	4083		
Qy	1936	lLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGl	1956		
Db	4084	GCCTTATCCC-----GTTCCGCTCTGGGTCCACTTCGCGCCCGCGCCAGGACCTC	4137		
Qy	1956	yHisAlaPheLeuAlaLysPro-----ProAlaArgSerGlyLeuGluProAl	1972		
Db	4138	C-----CTCGCCACCCCAACACCTTTTACCGCGTTTCAAGTACGCGATCGAC	4188		
Qy	1972	aSerSerProSerLysGlySerGluProArgProLeuValProPro-----Va	1988		
Db	4189	CTTCTCGCCCAAGACCGATCTCTCCAAAGCCCGCCGCCACCCCGCTCGACGCTG	4248		
Qy	1988	lSerGlyHisAlaThrIleAlaArgThrProAla-----	1999		
Db	4249	CTTCTTCAACAGCAGCTCCGCGGCTTCCATCTCCGTAACCCCTCTTCTCCCGCATC	4308		
Qy	2000	-----LysAsnLeuAlaProHisIleAla-----	2007		
Db	4309	ACGGTGGCGGCCACCCAGCCTTCTTCGCCCAACGCCCACTTCTTCAACCAACCTCCACGGC	4368		
Qy	2007	-----	2007		
Db	4369	CTCCGCCCAACCACTTCCCGGACAACTTCCCAACCGCCCACTTACCTGCGCGCTTCC	4428		
Qy	2008	-----SerProAspProAlaProProAlaSerAlaSerAspPr	2021		
Db	4429	CTCCCCCTACCAACGACCGAGTCTTACCCCACTCAGCCCTCCCGCACCGGTCTCTCCCC	4488		
Qy	2021	oHisArgGluLysThrGlnSerLysProPheSerIleGlnLeuLeuLeuArgSerLe	2041		

Db 4489 GCCTCCACCCGACTGCTCAGCGCCGACACCATCCCGAG----- 4531
 Qy 2041 uGlyTyHisGlySerTySerProGluGlyValGluProValSerPro-----Va 2059
 Db 4532 -----TCAACCCACCTTGTCCCGAGACCCGCTCCCTCCAGAGT 4575
 Qy 2059 lSerSerProSerLeuThrHis-----AspLysGlyLeuProLysH 2073
 Db 4576 GAACTCGCCCCACCCCTCCCATCCCAACCAAGACCTTCCCGACCTACCACTCCCGGCC 4635
 Qy 2073 isLeuGluGlu-----LeuAspLysSerHisLeuGluGlyG 2085
 Db 4636 ACCTTCGAGACCGCTTACCTCTGGAGTGGACGGCGAAGCCCTCGCCCGCACCTTCTTGGC 4695
 Qy 2085 luLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaHisLeuProH 2105
 Db 4696 GCCACCGACCGCTGGAGCTGGAGATCTT-----CTTCCGC 4731
 Qy 2105 is-----LeuArgProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrA 2123
 Db 4732 AACAACTGGTCCAAACAGTTCCTTCTCATCAACCGCCCGACACCGTAGCTGCAACCT 4791
 Qy 2123 laProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisLeuGluValI 2143
 Db 4792 CT-----CACCTCGT----- 4802
 Qy 2143 leThrGlnAspTyThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProL 2163
 Db 4803 -----CGCCCCACCCACACCAAAAGCAAGACCCACCTCTCCA- 4844
 Qy 2163 euTySerPheProGlyAlaSerCysProValLeuAspLeuArgPro----- 2179
 Db 4845 -----CGCTTCCCTCGCAAGCGTCTCGCTTCGCGACTCCACGAGCCCTACACCATC 4899
 Qy 2180 -----ProSerAspLeuTyLeuProPro-----ProAspHisGlyAlaProA 2194
 Db 4900 ACCGCCAAGGATCAAGCTTAGGCTACATCTCTTACCACCTCCCT-----CCAG 4947
 Qy 2194 laArgGlySerProHisGlyGlyGlyLysArgSerProGluProAsnLysThrSerV 2214
 Db 4948 CGCGCTTACTCGCTTCCCGAGCCGCTCCCTTCGACGACGCTCTCTTCGC-----G 5001
 Qy 2214 alLeuGlyGlyGlyGluAspGlyIleGluProValSerProGluGlyMetThrGluP 2234
 Db 5002 TCCTGTGATCGCGAGAACGA-----CTTCGCCACGCTCACCTCCAAACC 5046
 Qy 2234 roGlyHisSerArgSerAlaValTyProLeuLeuTyArgAspGlyGluGlnThrGluP 2254
 Db 5047 CAAGCCACCATCCAAAGCAAGCGCTTCGCTC-----CGACCCGGATT-GGCGTCACAC 5099
 Qy 2254 roSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerL 2274
 Db 5100 CTTCGTCGGCATCTCTCGAAGACCC---AGCAAGGTCAACGAGAACTCTCTTTTTCAC 5156
 Qy 2274 ysLeuThrGlu-----SerAsnSerAlaMetValLys-----SerLysL 2287
 Db 5157 CTCTTGGAAAGCCTCCGACCCCTCGCCCTCATCATGATTACTTGTCTCTGCTCGG 5216
 Qy 2287 ysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyAsnLys 2307
 Db 5217 CCCAGTAAAAAGATGATCAGCGCATCTCTCGACGCGGAGACCGCCCGCCCTCTCTCAT 5276
 Qy 2307 erGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrT 2327
 Db 5277 CCAGCGCGCGGACAGACCCCAACCACTTTCGAGTGTGCGAAGAACCTCTCACTCCCTC 5336
 Qy 2327 yrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeu--GluAlaIleI 2346
 Db 5337 CGTCCACTCGCCCATGACTACAGCGCTTCGACCAATCGCAGCACCGGTGAGCCGCTGT 5396
 Qy 2346 eArgLysAlaLeuMetGlyLysTyAspGlnTrpGluGlu--SerProProLeuSerAlaA 2366

Db 5397 GCTCGAGGCT-----TGAAGATGCGCGCGCCTTCATTCCTGA 5435
 Qy 2366 snAlaPheAsnPro-----LeuAsnAlaSerAlas 2376
 Db 5436 ACCTTTTCATCCTCCATCCGTCACGTCGACGACCAACATCGAATGCCAATTTGGCCCTCT 5495
 Qy 2376 erLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrS 2396
 Db 5496 CAGCTGATCGCATCACCGGTGAACCCAGGACCTAGCAGACACACCGACTACAACCT 5555
 Qy 2396 erProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysS 2416
 Db 5556 GGCCTCTCTTACACCCAGTACTCTCCACCCGACCCCGCTCTTGTGTCGCGGATGA 5615
 Qy 2416 erProAlaPro-----GlyLeuAlaSerGlyAspArgP 2427
 Db 5616 CTCTCTGTCGACGCGCTCCCCCATGAACCTTCTTGGCTGCTCG-----C 5666
 Qy 2427 roProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThr- 2446
 Db 5667 CCGCTCT-----TCGCTCTCAAGCCCAACCCGAGACCTCCCTTCGG 5711
 Qy 2447 -----AsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyRA 2465
 Db 5712 CTTTTCTGCGGTACTTCTCGCGCCCGCGCGCTCGCGCCCTTCCTTTT 5771
 Qy 2465 snProLeuIleMet-----ArgLeuGlnA 2473
 Db 5772 CGCAAGCTCGCATCGCTCGAGGACGCTCTCTCCCGAGAAATAGCTTTAGCT 5831
 Qy 2473 laGlyValMetAlaSerProProProGlyLeuProAlaGly---SerGlyProLeuA 2492
 Db 5832 CGCTAGTCTCGTCTCGTCTCGTCTCGTCTCGTCTCGTCTCGTCTCGTCTCGTCTCGT 5890
 Qy 2492 laGlyProHisHisAlaTrpAspGluProLysProLeuLeuCysSerGlnTyGluT 2512
 Db 5891 -----AACTGTCATCTACCATGTCAGCTGCTTTGACCTCATCTCCGCCCA 5936
 Qy 2512 hr 2512
 Db 5937 CG 5938

RESULT 72
 AP000547 123288 bp DNA linear PRI 10-JUN-2000
 LOCUS Homo sapiens genomic DNA, chromosome 22q11.2, Cat Eye Syndrome
 DEFINITION region, clone:KB67B5.
 ACCESSION AP000547 BA000006
 VERSION AP000547.1 GI:5931525
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 123288)
 AUTHORS Shimizu,N.
 TITLE Homo sapiens chromosome 22 clone BAC KB67B5 on 22q11.2
 JOURNAL Published Only in Database (1999)
 REFERENCE 2 (bases 1 to 123288)
 AUTHORS Shimizu,N.
 TITLE Direct Submission
 JOURNAL Submitted (22-SEP-1999) Nobuyoshi Shimizu, Keio University, School
 of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
 160-8582, Japan (E-mail:nshimizu@med.keio.ac.jp,
 tel:81-3-3351-2370, Fax:81-3-3351-2370)
 COMMENT This is a complete sequence of the insert of KB67B5 clone. The
 proximal adjacent clone is cl8E3 (Acc.#AP000546) with 4739-bp
 overlapping. The distal adjacent clone is KB7G2 (Acc.#AP000365)
 with 33343-bp overlapping.
 FEATURES
 source Location/Qualifiers
 1. 123288
 /organism="Homo sapiens"
 /mol_type="genomic DNA"

Db 74725 GCAGACTCCCTCAGTCAAGAAAAAGAAAAAGAAAAAACACCCAGCTG----- 74774
Qy 457 heLeuGluArgLysThrValAlaGluCysValLeuTyrTyrThrLysLys----- 475
Db 74775 -----GCCAATGGTGAAGACAGCTCTCCACTAAATATACAAAAAATAGTA 74823
Qy 476 -----AsnGluAsnTyrLys----- 480
Db 74824 TGGGCATGGTGCACTTGTCTCTAAATCCCACTCTCGGAGGCCCATGTAGAGGATCG 74883
Qy 481 -----SerLeuValArgArgSerTyrArgArgG 491
Db 74884 CTTGAACACAGAGGTGGAGCTGCAGCTAGCAAGTTTACACCATTCACCTCGCTG 74943
Qy 491 lyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 510
Db 74944 GCAAAACCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75003
Qy 511 ProMetProArgSerGln-----GluGluLysAspGluLysGluLysGlu 528
Db 75004 CCAGCTCGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75063
Qy 529 AlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAspLeuLysGlu 548
Db 75064 CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75096
Qy 549 LysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSerLysGly 568
Db 75097 GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75156
Qy 569 ArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgLysGlyArgLysGlyArgLys 588
Db 75157 GTGAAGCCAGCC-----AAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75213
Qy 589 GluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGlu 608
Db 75214 CCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75273
Qy 609 LeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeu 628
Db 75274 GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75324
Qy 629 GluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGln 648
Db 75325 -----GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75345
Qy 649 CysLysAsnPheTyrPheAsnTyrLysArgGlnAsnLeuAspGluLeuGlnGln 668
Db 75346 GCCACCCAG-----CCAGCCATACAGCCCAAGC 75372
Qy 669 HisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysLysLysLys 688
Db 75373 CAC-----CCAGTCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 75420
Qy 689 AlaAlaSerGluGluAlaPheProValValGluAspGluGluMetGluAlaSer 708
Db 75421 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75480
Qy 709 GlyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGly 728
Db 75481 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75525
Qy 729 AsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThr 748
Db 75526 -----CCAAACGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75576
Qy 749 GluSerIleProSer-----ProHisThrGluAlaAlaLys-----AspThrGlyGln 764
Db 75577 GACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75636
Qy 765 AsnGlyProLysPro-----ProAlaThrLeuGluValAlaAspGlyProPro 779
Db 75637 CAGGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75696

780 ProGlyProProThrProArgThrSerArgAlaProIleGluProThrProAla 799
75697 AAGCCACCACTAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75756
800 SerGluAlaThrGlyAlaProThrProProAlaProProSerProSerProSerAlaProPro 819
75757 AAGACAGCC--GGCCAGCCAAAGCGCGCAAG--CCACCCGCGCCAGCCAGCCAGCCAGCCAG 75812
820 ProValValProLysGluGluLysGluGluThrAlaAlaAlaProProValGluGlu 839
75813 CAC-----CCGGCAGCGAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75866
840 GlyGluGlnLysProAla-----AlaGluGluLeuAlaVal 853
75867 GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75926
854 AspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGluGly 873
75927 CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75980
874 ProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAla 893
75981 CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 76037
894 GluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGln 913
76038 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 76097
914 AspSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAsp 933
76098 AAGCCACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 76157
934 LysAsnArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAla 953
76158 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 76217
954 AsnAlaSer-----ProGlnLysProLeuAsp--LeuLysGlnLeuLysGlnAr 969
76218 CAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 76277
969 gAlaAlaAlaIleProProIleGlnValThrLysValHisGluProProArgGluAspAl 989
76278 CAAAGTGGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 76337
989 aAlaProThrLysProAlaProProAlaProProProProGlnAsnLeuGlnProGluSe 1009
76338 CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 76397
1009 rAspAlaProGlnPro--GlySerSerProArgGlyLysSerArgSerProAlaPro 1029
76398 T-----CCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 76451
1029 roAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeu-----ProGlyAspProp 1047
76452 CAAATCACCACCAAGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 76511
1047 roCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSerP 1067
76512 CGAAGCAACCAAG-----CCAGCC-----AGCCAGC 76538
1067 roHisAlaProAsp-----ProSerAlaPheSerTyrAlaProProGlyHisProLeuP 1085
76539 CAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 76586
1085 roLeuGlyLeuHisAspThrAlaArgProValLeuPro-----ArgProProThrIleSerA 1104
76587 ACTAAGCC--AGCCAGCCAAACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 76643
1104 snProProProLeu-----IleSerSerAlaLysHisProSerValLeuGluArgGlnI 1122
76644 GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 76703

Db	78538	CAGCCAAAGCCACCGCGCAGCAAGCACTTAAGCCACCGCGCCAGCCAGCCAGCCAAAGC	78597	Qy	2116	-SerSerProLeuLeuGlnThrAlaProGlyVallyGlyHisGln-----	2130
Qy	1793	-----SerSerGluArgGluArgAspArgAspa	1802	Db	79564	CACCCAGCCAGCCAGCAGCAGCCAAATCAGCCTGCAGCCAAAGCCAGCCAAAG	79623
Db	78598	CACCTGGGCAGCCAGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	78657	Qy	2131	-----ArgValValThrLeuAlaGlnHisIleSerG1	2141
Qy	1802	rgGluArgAspArgAspArgGluArgGluArgSerIleLeuThrThrThrValG	1822	Db	79624	CCACCCAGCCAGCCAGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	79683
Db	78658	CTGCCAGACAGCAATCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	78717	Qy	2141	uValleThrGlnAspTyrThrArgHisHisProGlnGln-LeuSerAlaProLeu	2159
Qy	1822	luHisAlaProIleTyrArgProGlyThrGluGlnSerSer-----	1835	Db	79684	GCCACCCAGCCAGCCAGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	79743
Db	78718	CCACCCAGCCAGCC-----AAACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	78774	Qy	2160	--ProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgp	2179
Qy	1836	-----GlySerSerGlySerGlyGlyGlyGlyGlyGlySerSers	1849	Db	79744	AGCCATCCAGCTGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	79797
Db	78775	CAAGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	78834	Qy	2179	roProSerAspLeuTyrLeuProProProPheHisGlyAlaProAlaArgGlySerProH	2199
Qy	1849	erArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnA	1869	Db	79798	CACCCCAAG-----CCAGCCAAAGTACCCCGCCAGCCAGCCAGCCAGCCAGCC	79845
Db	78835	CCAGCCATCCAGCCAGCC-----AGCCAGCGCGCCAGCCAGCCAGCCAGCCAGCC	78891	Qy	2199	isSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValleuGlyGlyG	2219
Qy	1869	spAlaLeuGlnArgProSerValleuHisenThrGlyMetLysGlyIleIleThrA	1889	Db	79846	ATGCC-----AGCCAAAGCCAGCCAGCCATCCAGCCAGCC-----	79880
Db	78892	CAGCCAAAGCCACCCAAAGCC-----CCCGCCAGCCAGCCAGCCATCCAGCCCA	78936	Qy	2219	luAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHis	2236
Qy	1889	la-----ValGluProSerLysProThrValLeuArgSerThrS	1904	Db	79881	-----AAGCCACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	79929
Db	78937	GCCAAAGTCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	78996	Qy	2237	-----SerArgSerAla-----ValTyrProLeuLeuTyrArgAsp-GlyGluGlnThr	2252
Qy	1904	erSerProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyT	1924	Db	79930	GCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	79989
Db	78997	GCCAGCCAGCCAAACCCAGCC-----AAGCCAGCCAGCC-----	79028	Qy	2253	GluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhe	2272
Qy	1924	hrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProA	1944	Db	79990	CAGCCCA-----GCAAGCCACCCCGCCAGCCAGCCAGCCAGCCAGCCAGCC	80037
Db	79029	-----CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	79059	Qy	2273	SerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnLeuLeuAsnLys	2292
Qy	1944	rgValAlaArgProGluArgProArg-AlaAspThrGlyHisAlaPheLeuAlaLysPro	1963	Db	80038	AGCCATCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	80061
Db	79060	AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	79113	Qy	2293	LysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGlu	2312
Qy	1964	ProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgPro	1983	Db	80062	-----CAACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	80106
Db	79114	CCGCGCCAGCCAGCC-----AAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	79170	Qy	2313	IlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVal	2332
Qy	1984	LeuValProProValSer-----GlyHisAlaThrIleAlaArgThrProAlaLysAsn	2001	Db	80107	-----AGTCTGCCAGCC-----AGCCAAAGCCAGCC	80130
Db	79171	CCAAGTCACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	79230	Qy	2333	GlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGly	2352
Qy	2002	LeuAlaProHisAlaSerProAspProProAlaProProAlaSerAlaSerAspPro	2021	Db	80131	CAAGACAGCAGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	80166
Db	79231	CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	79290	Qy	2353	LysTyrAspGlnTyrGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsn	2372
Qy	2022	HisArgGluLysThr-----GlnSerLysProPheSerIleGlnGluLeuLeuArgSer	2040	Db	80167	-----AGCCAGCCAAACCCAGCC-----CAAGCCAGCCAGCCAGCCAGCC	80202
Db	79291	GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	79326	Qy	2373	AlaSerAlaSerLeuProAla-----AlaMetProIleThrAlaAlaAspGlyArgSer	2390
Qy	2041	LeuGlyTyrHisGlySerTyrSerProGluGlyValGluProValSerProValSer	2060	Db	80203	-----CCAGCCAAAGTATCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	80259
Db	79327	AAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	79383	Qy	2391	AspHisThrLeuThrSer-----ProGlyGlyGlyLys-----AlaLysValSerGlyArg	2408
Qy	2061	SerProSer-----LeuThrHisAspLysGlyLeuProLysHisLeuGlu	2075	Db	80260	CAAGTCAGCCAAACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	80319
Db	79384	CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	79443	Qy	2409	ProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProPro	2428
Qy	2076	GluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLys	2095	Db	80320	CCGCGCCAGCCAAAGTATCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	80373
Db	79444	CAGTCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	79503	Qy	2429	SerValSerSerValHisSerGlu-----GlyAspCysAsnArgArgThrProLeuThrAsn	2447
Qy	2096	LeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuGluSerGlnProSer	2115	Db	80374	GGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	80421
Db	79504	AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	79563				


```
/codon_start=1
/transl_table=11
/product="Stah"
/protein_id="AAM80527.1"
/db_xref="GI:21911404"
/translation="MAGLAVFAVAFVAFARLAKRTIROPVNGEIIAGLALGPSILGL
LPQDVEALFPADARPVQLSOLGVLVFMFGVIGYELDFALLRGAGSRVTLVSLSSMA
LPMALGAVLAVLFDYKVSQSSDGLVAPLPLGVAMSIITAPPVLARIAGNQRT
RICMALACAVQDFLAWCVLAVAVIATASGLMPLARMLVLESGLFLGVLLVVRPGL
RRLALPERRAVGSSLIHALLVTLGLVSAWATHEIGHAVFGAFAGVAVPHRHIEVQ
APOVERLEQTSIFLLPFTVTGLSDVDVGLGRTGLIMTVAVLVACTGKTVGAVS
ARTIGAKSERSILGVILNARGITELVILNVGLGLDRLSRLFSAMVVALVTFTMTG
PLIHGHEELRLVFPVAKNSV"
/complement(2485..4071)
/notes="putative nonheme iron dioxygenase"
```

CDS

```
/codon_start=1
/transl_table=11
/product="Stah"
/protein_id="AAM80528.1"
/db_xref="GI:21911405"
/translation="MTEEQFLRSDTIIEPLVDRFVAQLYTTAPVSGTWNLAFRQIPM
LESYLPDYPHFGSKPKRGFFVGIIEKKRGEVEALIAAIRDRADMIRFAEAI
BAEKIVREERTYDLRLSLYKPLSELAGLEVETDTCNQASHLEFLRAYQSKAYTED
RQSVLSLETGIERPFVMSLPRLSSDLVLELDFPRHPGIEELFLSRITPTVAALRE
ALEGDRARAQSLRLLVAEPLAADRHVTAETIRYKWHACILLQTPVSLITDPTIS
ADTSATRYTDDLDHIDIVITHGSHDLVAETLLQLRGVGTVPVTRSRGNLGD
PSLALFKLGLVIEVDDEVPFGGKIIVATFFGEGHADLDIRAKSTIWISIGKS
VYVAGDSGIDPVLVYRVRHLGRVDIAFLGMECDGAPMNKYPQMLTQPIPKMSDT
RKLGSNAEQASATVTELGAEVVIYANGESWLGHVMTSYNEDSYQLQQTIAEFEEW
CTRGQVKAGHLLNRAEHWHA"
/complement(4151..4963)
/notes="sulfoltransferase"
```

CDS

```
/codon_start=1
/transl_table=11
/product="Stal"
/protein_id="AAM80529.1"
/db_xref="GI:21911406"
/translation="MNGCMWIASYPKAGGHWLRCLMTSYVTGEPVTPWPGIOAGVPHL
EGLLRDEAPADDEQVLLAHTHTADRPULRYFRESTAKVCLVCLNPRDMLSLMRM
KIPPEDEACRKLITFADDEGSTRVIRIWAGEGSPENIRSWTDSVHESFPNAAVLA
VRYEDRLKPEGLMKVVDVLELQGRDGVADAVANCETLERNMEERSKLLGLETGL
MYRGKQLPFVFGQGRKSLKFMGDDIEKAYADLLHGETDFAHYARLYGAE"
/complement(4960..6267)
/notes="Nonheme halogenase"
```

CDS

```
/codon_start=1
/transl_table=11
/product="Stak"
/protein_id="AAM80530.1"
/db_xref="GI:21911407"
/translation="MSAETFDVVAGGPGAGSTAASLVAMQGHVLLVEKEVILPHRV
GESLLPSTVHGCMRLGVTDELAETDPVKEGATFONGRAEPWAFSGASRLTDTAT
SFAYQDARDEILKNARKKGVVREGCVTAIVREGDVTGLRHTDPDGGHEVGT
ARFVTDASGHESPLSASVGGTRDRAGSRGLALSGYFGNGCKRULPERPSGDLSTVD
GWFVYPLFGGLTVSGVAVSEERAEKIQDDREAFKALTAECFAISDLSEGMKRVTA
GKYGLRVEDYVYHRTTFRPQGMILLVGDAAACVDPLFSGVHLATYALLAARSINS
VLAVGDSKTLATEPEARYRELGAFFSELLTFTMTGEEHPESEFVDVAAVSSGE
WALAQQGGAAPCEPMRGGVLAIVSDGKWLPAASKEREDAR"
/complement(6312..7490)
/notes="P450-related oxidase"
```

CDS

```
/codon_start=1
/transl_table=11
/product="Staj"
/protein_id="AAM80531.1"
/db_xref="GI:21911408"
/translation="MRRITCDPHEDMPALRAHGLIRIEGNASQMSDSTYVQWQMGVD
VYVKTLGHNFTLRITLTDQAPLQSGEGVSVPPELAGQISYIDPEPTRLRMLTPEP
TVRRRLRRELEGGIEEHLDALEGAPPADLQVLPADPVGGETLCMLLVGPRDRNE
TVRRRVNVDLSRGYKARAADSAFNRYLMTLITQRKDPDDEGLMLVREHGDRITD
EELKGVCTALIGGVSVAGMIGFVALLLEHPDORLLFGSREEARLNNELLRFLS
AVQQTPTPMANRDVVVEQLIKAGEVYLCSLIMANRDGLTSDSHLLDANREPLPHVA
FGHGHHCIGAAVAAVULKITQSLWRRFPRLSLAVPAGEVKFNFAFIDSPDLRPLVTW"
/complement(7534..9072)
```

CDS

```
/note="Nonheme halogenase"
/codon_start=1
/transl_table=11
/product="Stah"
/protein_id="AAM80532.1"
/db_xref="GI:21911409"
/translation="MDGLSGDPLPSTIAKDSIEGWEMSVETFDVVVAGGPGGSTVA
ALVAMQHRVLLLEKEVPFPRYOIGESLLPSTVHGVCRLLGVTDELAALKFVKRGGTF
RWGARPWFESFVSRMTGPTSFAYQVERARDEILLNNAKSGVTVREGCVTAIV
VEGRDVTGLRITSDSGEHEVGARFVIDAGSKSRSLHVASVGGTREYSEPRSLALFG
YFENGKRLPEPSYGNILSVAFSDSFWFWIPLSDTLTSVGVVVRREMAEKLOGDEKAY
NALLAECPLTIAEYLSADARVTYGNLGRVKDYSYHQTFKRFWRCEMLVGDAAFCFDP
VFSGVHLATYSALLAARSINSVLAGDLDEKTADEFEGRYRRYGVYFVEFLVSFYQM
NNESYFWHAKVKTNNHETEFESFVILGVASGETALASRLAERSAEAFAGVADQM
AGDDGDMDVPLFKSHVVKVQVMEGQVQMKAVLGEDAAPERPVPFGGLVSPDGKMWL
PYRG"
/complement(9077..10273)
/notes="P450-related oxidase"
/codon_start=1
/transl_table=11
/product="Stah"
/protein_id="AAM80533.1"
/db_xref="GI:21911410"
/translation="MSGDDRRPIHTLRQGFDPADLRAAGELTVRLSGGADAEHTWL
ATGHDVVRQVLGDHTRFSTRRRFDRNDEIGKGVFRPRELVNLMMDYDPPETRLRRL
LAPGTHRKIRRMAPYIQIVTERLDEMERESEPADLIELFADEVPGLVCELGVPR
DDRMFLOLCHRHLDASLSGRRRAAGAEAFSYLVTVWABRKQDPDGLIGMVAEHG
DTVTDEELRGVCOMLAGDDNISQMLGLVALLRNPQIAALRGDVPAAERAVDEL
IKYLVTPAFTPTIEDSTYGDQVIKAGETVLCSLPTANRDPALLPADRLDVTREA
YHVFAPGHHVHCLGAALRLRIAYTALWRRFPDLADPDGATEFRLSTPAPGIS
RLMVTW"
/complement(10263..11417)
/notes="P450-related oxidase"
/codon_start=1
/transl_table=11
/product="Stag"
/protein_id="AAM80534.1"
/db_xref="GI:21911411"
/translation="MALPLHRRHRLDPVEPHDIQNEGPLHEYTEPMGDKRKOWLV
TGHEVRDITLADPERFSSMRPVDDEADRALLEGILQAYDPPDHTLRRTVAPASRR
MERLPRIEIVEECLDDLEVDGSPVDFVRYAAMPILALICEFLEDPDRDRAELSRM
IRESRESRLPROTSGMGVNNYTKLAARKLDPGEGMIGVIREHGAEDVDELAG
LAEGNIMAAEQAAQAVALLVTHPDQALLAREHPELVDPGATEELLRHASIIVEAP
APRVLEDVSVAGRIDRAGDVLTCMMANVRPQGEHFOITRENPKHMAFGYGIHCHLG
APLRLQLRVALPAVLRFPRLVPEEDLRFKPGRPAPPAVEELPVEW"
/complement(11438..12613)
/notes="P450-related oxidase"
/codon_start=1
/transl_table=11
/product="Staf"
/protein_id="AAM80535.1"
/db_xref="GI:21911412"
/translation="MFEINVVRASQLHRRDRDPVPELHSLMKEGGLTVLGTEDSTE
GFWHMRQKDERVEQIYVDSLDTIESLGAGTVPFGFLTAQDTPETRLRMMVVP
ATSHRRMQFRPEVEQIYVDSLDTIESLGAGTVPFGFLTAQDTPETRLRMMVVP
ADUARILLASRTDRSKRTAGNKEMTYMKQHVAGSRSGDGLRGIIVRENGDAIT
DAELTGVAAFGMAAADQVRLAAGLWLVQAFQALLRKPEFVPEWLEDETVL
TTDEKTHPRVATQDVRIGQVKGADVTVCSLAANRPNYPGADEFDITREKAEHLA
FGHGIHCHLGRMAELMFKVSI PALAHRFPFTRLADPQREITLGPFPFDFVEALLLDW"
```

Alignment Scores:			
Pred. No.:	2.54e-05	Length:	69301
Score:	689.00	Matches:	674
Percent Similarity:	32.25%	Conservative:	301
Best Local Similarity:	22.30%	Mismatches:	1181
Query Match:	5.21%	Indels:	873
DB:	1	Gaps:	144
US-09-522-753-5 (1-2517) x STU02965 (1-69301)			
Oy	14	AlaThrGluProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArgThr	33
Db	31302	GCTCGTGAACCGGCCCCACCGTCGGGACGGATC---CATCCC-----GCCACC	31349

```
QY 34 HisThrAspValGlyLeuLeuGluTyrGlnHisHis-----SerArgAsp 48
Db 31350 ATGACGCGCGTGTCTCCAGCTCCACACGACACCACTGCGCTCTCCGCCAACGCGCCAC 31409
QY 49 TyrAla-----SerHisLeuSerProGlySerIleLeuGlnProGln--- 62
Db 31410 CGCGCCGCCAGGCGCGTGCAGCAGCAGCTCTCGTGACACCACTGAGTGA-AGCGCCAGGTGA 31468
QY 63 -----ArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsn 76
Db 31469 CCTGCTCCGCGAGCGCACCGCGCGGTCTCTTCGCAAGGTCCATGATC-----GGCGTC 31522
QY 77 GluArgSerGlnGluHisLeuArgProGluSerHisSerTyrLeuPro---GluLeu 95
Db 31523 CCGCGGCGAGCGGTCCGCGGTCCGCCA-----GCCGTGGTCTCTCTCCGTCGA 31573
QY 96 GlyLysSerGluMetGluPheIleGluSerLysArgProArgLeuGluLeuLeuProAsp 115
Db 31574 GGAGCGCGTCCACCCCGTCACTTCGCGCAGCGGTTCGCGGAGGTGCG-----CCTGCT 31627
QY 116 ProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeu 135
Db 31628 CCGCAATGCGGCG-----CCATCGCGGAACGAGAGCTCA-----31666
QY 136 ThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProProSerProProHis 155
Db 31667 -----CCGCGTCCGAGCGAGCGTCTCCCTCGCGCGCGACCGCTCAC 31708
QY 156 ThrAspProLufLeuGluLeuVal-----ProProArgLeu 167
Db 31709 AGGACGCGCTCAGTCCGCGCAGCAGGATCTCCACGACCTCGCGTCCGACACCGCGGGT 31768
QY 168 SerLysGlnGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGlu 187
Db 31769 GCGCCACCGACAGCGCGTCCGACAGCGGTCCGCGT-----31810
QY 188 GlnGlnIleSerLysLeuLysLysLysGlnGlnGlnLeuGluGluAlaAlaLysPro 207
Db 31811 -----CCA 31813
QY 208 ProGluProLufLysProValSerPro-----ProProIleGlu 220
Db 31814 CCCACCGCAGCGACCATCAGCGCGCGCGAGGATCCAGCGCGCGCGCTCCCGAG 31873
QY 221 -SerLysHisArgSerLeuValGlnIleIle-----TyrAspGluAs 234
Db 31874 CGCACCGCTCCGCGACGCTGCTCGAGGTACCGGTGTCGCGTCCGCTCGACCTCGATCAGAT 31933
QY 234 nArg-----LysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnVa 252
Db 31934 CGCGGGGTCCACCGCGCCCTCGCGGACAGCACCATCCCGGGGCGTCCCGCCCGACGG 31993
QY 252 lGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysII 272
Db 31994 TCGGCTCCCGCAACATGCTCGTGGCGTGCATCATCCAGCTCAGCGCGCGCGCGCGGTGT 32053
QY 272 eAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArg-----286
Db 32054 CCACCGACAGTACCGGGCGCGCGAGCGGTCCACCCACCGCGCGCGCGCGCGCGCCCGAGA 32113
QY 287 -ArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMe 306
Db 32114 CCTTCATCGGCGTCCGGCGAGCGTCTCCGCGTCTCGGGGCGGTTCGCGCGACCGCG 32173
QY 306 tGluAlaLeuGlu-----LysLysValGluAr 315
Db 32174 CGTCCAGTCCAGGATCCCGCGGATTCGCGCGGGTCTCTGTGTCGAACACCTCAGCG 32233
QY 315 gIleGluAsnAsn-ProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluL 335
Db 32234 CGCGAAGACCATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 32293
```

```
QY 335 ysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArg--- 353
Db 32294 CCGCG---CCCAACTCGAAGAAGTGTTCATCCGCGCGCGACCCGATCCAGGTTCGAGAACCT 32350
QY 354 -----ValGlyGlnArg-----357
Db 32351 CGCGGAACAGCGCGCACAGCGCGCTCGGCTCGGTGACCGGTTCGCGCGCGGTGACGC 32410
QY 358 -----GlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValS 373
Db 32411 GTCGCGCGAAGTCTGGGTCCGCG-----AGCGCGCGCGGTCCACCTT-CGCGTTG 32460
QY 373 erGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGlyLysGlnMetArgGlnLeuA 393
Db 32461 GCCGT-----CACCGCAGCGCGTTCGAGGACAGCACCGCGCGC---32498
QY 393 laValIleProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetA 413
Db 32499 -----CGGCACCATGTAGCCGCGCGCGCGCGCTG 32526
QY 413 snGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrps 433
Db 32527 GCACGCTCTCGCGCAGCCCTTCGGGTG-----GACGTGCGCGCGCGGACACACATAG 32580
QY 433 erGluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyL 453
Db 32581 CCGATCAGGCGTCTTCACGTGCCACACCGCGCTGGTTCGACGCGCGCTCAGCC 32640
QY 453 euIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrLeuT 473
Db 32641 AGCACGCGCTCGATCTCGCGCGCTCCACACGCTAGCCGCGGATCTTCACGTGTTGTCG 32700
QY 473 hrLysLysAsnGluAenTyrLysSerLeuValArgSerTyrArgArgGlyLys 493
Db 32701 GCAGCGCGCGGTACACCGCTCTCGCTTCGCTCC-AGTACGCGACGTCCGCGCTCCG 32759
QY 493 erGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnPro--- 511
Db 32760 GTACATGCGCTCACCGGAGCGGAGCGGTCCGCGACGACCGTTCGCGCTCAGCGCGG 32819
QY 512 -----MetProArgSerSerGlnGluGluLysAspGluLysGlu 524
Db 32820 ACGGTCCAGATAGCCGTGCGCCACCCCGCGCGCGC-----GATGTAGAG 32864
QY 525 LysGluLysGluAlaGluLysGluGluLysProGluValGluAsnAspLysGluAsp 544
Db 32865 ATCGCGCTCGACCGCGCGCAACGAGCAGGACCGAAGACATCCAGAC-----32912
QY 545 LeuLeuLysGluLysThrAspThrSerGlyGluAspAsnAspGluLysGluAlaVal 564
Db 32913 ---GTACAGCGCGCGCGCGCGAGTGGCGCGCGATCCGCGAGGACCGATCCAGCGCGTC 32969
QY 565 AlaSerLysGlyArgLysThrAlaAsnSerGlnGly-----ArgArgLysGlyArgIle 582
Db 32970 GCTGGGTTCGATCGCTTCAGGTTCGACACAGGCGTGTCTCGGTTCGCGCGCTGCTGTG 33029
QY 583 -ThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSer-- 601
Db 33030 CCAGACGACGCTCGGGGCGAGCGCGCGCGCGCGCGCTCCACCTCCAGCGCGCGCACAC 33089
QY 602 -----AlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluG 618
Db 33090 GTCGCGCGCGCGCGCGAGCATCGCGAGCCCCGAGACGACTCCGG-CGACTCTCTGGACCA 33148
QY 618 uGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerIleAla 638
Db 33149 GCCTCGGAACACTACCGCGCGTGAAGTT-----CACGGA-----GGTCAGCGCTCTG 33196
QY 638 aArgMetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLy 658
Db 33197 CGACATGAGCGCGAGCGCGCTC-----33218
QY 658 sArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAs 678
```


Db	33219	-CGCGTCACACCGCCCGGCTCGCGAGCAGCACCG--CGCACCCGACAGAGCGCA	33274
QY	678	n-----AlaArgArg-----Ly	682
Db	33275	CCAGAGTGTACAGGAGCGGTGGAAGGTGCGCGCGGTGCATCAGCAGGTGTGCG	33334
QY	682	sLysLysLysAlaProAlaAlaAlaSerGluGlu-AlaAlaPheProProValValGluA	702
Db	33335	CGGTCCGACGCCCGCCCGGTCTCCCGAGAGCGCGACCGCGCGGTGCGCAGCG	33394
QY	702	spGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGluAag	722
Db	33395	CGAGCGCTTG-----GGTGTCCCGTGCATCCGAGGTGTACATCAGTACGCCA	33445
QY	722	luAlaLeuHisAlaSerGly---AsnGluValProArgGlyGluCysSerGlyProAlar	741
Db	33446	GTGTCTCCGCGCCGACCGGACCGAGCGAGTGC-----TCACCGTCA	33487
QY	741	hrValAsnAsn-----SerSerAspThrGluSerIleProSerP	754
Db	33488	CGGCATCGTCCATCGTACGAGCACGTGGCGCGAGGCGCACACGCGCGGTCCCGCAC	33547
QY	754	roHisThrGluAlaAlaLysAspThrGly-----GlnAsnGly-ProLysProProAla	771
Db	33548	CGCACACCGCGCGCGACCGCGCGTCCGACATCATGAACGGACCCCGTCCCGCGGT	33607
QY	772	ThrLeuGlyAlaAspGlyProProProGlyProProThr-----ProPro	786
Db	33608	AGTCCGGTGCACGGGACAGNACCGCTCCCGCTCCACACCCGAGCGCATCGCGCA	33667
QY	787	ArgArgThrSerArgAlaPro-----IleGluProThr	797
Db	33668	CTACGTCGCGGAACGCTCCAGCAGGACACCGACCGGTCCACACGTCGACGCGCCGGT	33727
QY	798	ProAlaSerGluAlaThrGlyAlaProThrProProAlaPro-----	812
Db	33728	CCTTCAGTACCGCGCGACCGCGCGCGTCTCGTCTCCAGCTCGCGAACGACCGTTC	33787
QY	813	-----ProSerProSerAla-----	817
Db	33788	GTCCCCCGCGTATCGCGCTCGCGTCCGCGCGCTCCACCGTCCGCGAGGAGAGAT	33847
QY	818	ProProProValValProLysGluGluGluGluThrAlaAlaAlaProProVal	837
Db	33848	CCACCGCGCGCGCGCGCGTCTCGTCTCCAGCTCGCGAACGACCGTTC	33907
QY	838	Glu-----GluGlyGluGluGlnLysProProAlaAlaGluGlu	850
Db	33908	CGCGTTCGCGTCCGCGCAGCGGTTCAGCGCGCGCGCGCGCGCTCCCGCTCCCGATCA	33967
QY	851	LeuAlaValAspThrGlyLysAla-----GluGluPro-----	861
Db	33968	TCTGTCTCCAGGACGCGGACGAGCGCGTGGAGAGCGCTCCGCGACTCCGCGTCGACGA	34027
QY	862	-----ValLysSerGluCysThrGluGluAla-----GluGluGly	873
Db	34028	TCTCCGCTGTGTGTCAGTTCGATCTCGATCTCTCACCGCGCATGCGCCGAGATGA	34087
QY	874	ProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAla	893
Db	34088	TCGCGTAGGGG-----ACGCTGTGTGGAACCTCTT-----	34117
QY	894	GluLysLysGluGlyGlySerGlyArgAlaThrAlaLysSerSerGlyAlaProGln	913
Db	34118	-----GCAGGGTGAAGCGCGA---GAGCGCGCGGGAGCTTCGAGCCCGCGCG	34162
QY	914	AspSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyAsp	933
Db	34163	GATCGACGTAGTCTCGAACA-----	34183
QY	934	LysAsnArgLeu-LeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAl	953

Db	34184	---CCACGAGCGTGTCCAGACCCCGCGCGAGCTTCTGTGATCTCGGGAGCCCGAGAT	34240
QY	953	aAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLysGln-----Ar	969
Db	34241	GCTGTGCGCGATCAGCGTGGACTCGCGCTCCTCAGCTCCGTGAGCGTCTCCATGACCG	34300
QY	969	gAlaAlaAlaIleProProIleGlnValThrLysValHisGluProProArgGluAspAl	989
Db	34301	TTCGCGCGCGTCCAGCG-----CACGCGCACCGGACCGGTGTGTGATGAACA	34348
QY	989	aAlaProThrLysProAlaProAlaProProProGlnAsnLeuGlnProGluSe	1009
Db	34349	-GCCCA-----CCATGCCCTCCAGCGCGGAGTCCCGCGCGCGCGAGACCGTCC	34401
QY	1009	rAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaProPr	1029
Db	34402	GCGCGAACATCATCTCTGCGCGCGCGCAGTCCGGCCAGCACGACGCGCCCGCC	34461
QY	1029	oAlaAspLysGluAlaPheAlaAlaGluAlaGln---LysLeuPro-----	1043
Db	34462	TGCACACCGTGTTCATCGTCAGCCCATGCGCGCGCGGCGAACTCCGTGAGGCGGAGTG	34521
QY	1044	-----GlyAspProProCysTrpThrSerGlyLeuProPheProValPr	1058
Db	34522	GAGCGCGGTGAATGCGGATCACCTCC-----GGACCGGCGAGTTCGCGTGC	34569
QY	1058	oProArgGluValIleLysAla---SerPro-----HisAlaProAspProSe	1073
Db	34570	CGCGCGCGCGTTCGAGATGTGAGCGCGATGATCGTCCGACCGCGCAGTTCGCG	34629
QY	1073	rAlaPheSerTyraProProGly---HisProLeuProLeuGlyLeuHisAspThrAl	1092
Db	34630	CGCCATG-----CGCGCGCGCGCGTCTTGTCT-----GC	34662
QY	1092	aArgProValLeuPro-----ArgProProThrIleSerAsnProProPr	1107
Db	34663	CTCCCGACCATGCCAGATGTTTCGGTAGGAGCGCGCGACCGCAGAC-----	34711
QY	1107	oLeuIleSerSerAlaLysHisProSerValLeuGluArg-GlnIleGlyAlaIleSerG	1127
Db	34712	-----CCGACACGACTCCGCGCGCGGTAGATCTCGAGGAGT	34749
QY	1127	lnGlyMetSerValGlnLeuHisValProTySerGluHisAlaLysAlaProValGlyP	1147
Db	34750	TGCGACTCGAGATCCCGCGCACCGCTCCATGAGCAGGTGGTGGTCTGCTGTCACG	34809
QY	1147	roValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyV	1167
Db	34810	AGACA-----GTGCGCTTCGACCGCGCGGATGAGCGTACGCG-----	34850
QY	1167	alLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeu-----	1184
Db	34851	-----CAGCAGCGGAGCGCGCT---CAGTCTGAACCCCTCGCGCGCTCTTCTGC	34899
QY	1185	-----GlyValProThrAlaGlnGluAlaSerVal-----	1194
Db	34900	GCGATCAGCTGACCTCGGCCATCGCGTCTCTCGCTAGGCGCGACAGTCCGCTCT	34959
QY	1195	-----LeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleT	1209
Db	34960	GTCGAGGAAGTTGGACATCCCGAGCAATGAGTGGACCGTCTCGCGGACTTCCGTCGG	35019
QY	1209	hrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyArg-----	1226
Db	35020	TGGAAGGACCGCGGAGCGTGGGTGCGGTGCGGTGCGAGGAGGCGCTCCCAAGACGTCCG	35079
QY	1226	-----	1226
Db	35080	CGCTCGGTGTCCAGCGCGTGTGATGATCTGCGTCTGCGATGACGTAGATGTCGCGG	35139
QY	1227	-----GlySerIleThrHis-Gly-----	1232
Db	35140	GCGCGTCTCGAAGGAGGAGTGTGATGAGCATTCCTCTCTGATCGGGGACGCGCCAG	35199

```
Qy 1233 -----ThrProAlaAspValLeuTyrLysGlyThrIleThrArgIle 1247
Db 35200 ATCTCGGCGAGCGCGATCCAGGAGCGTAGTCGTTGGGTACCGACTC--- 35256
Qy 1248 GlyGluAspSerPro-----SerArgLeuAspArgGlyArg 1259
Db 35257 ---CGCGAAGTCGTCGCTATGGCTTCAAACCCCTCGATCTCGTCTGATCGAGGTCGA 35313
Qy 1260 GluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSer 1279
Db 35314 GAA----- 35316
Qy 1280 TyrGluGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGly 1299
Db 35317 -----GGTCGAAGTCGAGCGGTGTGCCCGCGCGAGGAGTCTGCCCGCAGGT 35367
Qy 1300 ProProHisGluThrAlaAlaPro-----LysArgThrTyrAspMetMetGluGly 1316
Db 35368 GGGCCAGCGCGGACAGCGGTGTCCGCCACATCCGACCGAACCGTTCTGATCTCGGCTCTTT 35427
Qy 1317 ArgValGlyArgAlaIleSerSer-----AlaSerIleGluGly-LeuMetGlyArgAl 1334
Db 35428 CGAGGAGCGCGCTGCCAGCTCCCATGAGCTTCAGTTCCGGGCCCTG-----CGCA 35481
Qy 1334 alleProProGluArgHisSerProHisHisLeuLysGluGlnHis---HisIleArgG1 1353
Db 35482 TGTCGCGGACGACACCATGTCAGCTGCGGCGATGCGGCGAGTCCAGG 35541
Qy 1353 Yser----- 1359
Db 35542 AACCGCGCATGTCAACCGCGCTGTGCCACTCCGCGCATCGTGTCTGCTGCCCGCAGCGCA 35601
Qy 1359 eProArgSerTyrValGluAlaGln----- 1372
Db 35602 ACCGCGCCATGTAGTTGAATCCGATCCAGGCGGACGCGATGGCTCCAGCAGCGGCCCCG 35661
Qy 1372 gArgGluAlaLysLeuLysArgGluGlyThr-ProProProProProSerArgA 1392
Db 35662 TCGCGGGTTCAGATCGGCAACAGCGCGTGCACCGAGCCGTATCATCGACCGCGCGGTG 35721
Qy 1392 spLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisG 1412
Db 35722 ACTGCTCTTCCACGCGCTTACACAGATCC-----CCGACGCCA 35760
Qy 1412 LuGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluG 1432
Db 35761 GGTGATTCCTCCGAGACGTCCAGGCGCACCGGGTGACAGCTGGTGAACACGACCGCGTCC 35820
Qy 1432 LuLeuArgHisThrProGluLeuProLeuAlaProArgPro---LeuLysGluGlySerI 1451
Db 35821 GGGACAGATCAGCCCG-----CCGACCGGTGCACCGCGGTGACTCTCCACGTCCACCA 35874
Qy 1451 leThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysH 1471
Db 35875 GCACACACGCGCGCGGAGCGCGCCACCGCGCGCGCCAAACCGCGGAGC----- 35927
Qy 1471 isAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisPro--- 1489
Db 35928 -----AGTATCTGTGACCCCGCAGAGAACACGCCCGGGAACAGTCTCCACGA 35976
Qy 1490 --LeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeu 1509
Db 35977 GTCCGAGCGCTCGGCTCCCGCATATGACCCAGCAGCTGTGAGCCACACACTCCCGGTGCTCC 36036
Qy 1509 ysserArgProGlyThrAlaSerSer-----SerGlyGlyS 1521
Db 36037 CTTGCGCGCGCTGACCGCGCGGTTCACCTCGGTTCAGGATCGCTGCGCAGCGCGGAGTT 36096
Qy 1521 erIleAlaArgGlyAlaProValIle-ValProGluLeuGlyLysProArgGlnSerPro 1540
Db 36097 CGCGGACCGCGTCCGCGGAGTCCGCCACTCCACAGCAGGCTGCGCCACCGCGGGAACG 36156
Qy 1541 LeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProVal 1560
Db 36157 -----GCACCGCCACCGGCTCCAGCGCAGGGGTCCGTCCCGCAGGACCGCTCGC 36207
Qy 1561 Thr---MetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerSerAla 1579
Db 36208 ACCCGCGCGCGATCGGCACCA-----GAATCCGCGCAGGACA 36246
Qy 1580 SerGluAspArgLysLeuThrSerThrProArgGluIleAla----- 1593
Db 36247 CGCGGTCCGAGCGCATGTGTGCACCGCATACCAATCGGCCACCGCTTCCGGCCCGG 36306
Qy 1594 -----LysSerProHisSerThrValProGluHis 1603
Db 36307 CGTCCACCCACACCGCTGCACCATCAGCGCGCTCTCCGGTCCAGCGCTCCACCGCT 36366
Qy 1604 ---HisProHisProIleSerProTyrGluHisLeuLeuArgGlyVal-----SerGly 1620
Db 36367 CCGCACCGCGCGCTCCACGCGCTCTCCAGCAGCGCTCTCGTCCGAGCCCGCACCGCTCCGGA 36426
Qy 1621 ValAsp-----LeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIlePro 1638
Db 36427 CCGACCGCGCGGTGTCCAGGAGCCAGTTCACCTACGA-----TCAGGCGTCCC 36477
Qy 1639 ArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsn 1658
Db 36478 ---GCTCCACCGCTGCGCAACATGTGTGGGTGTGCACGAGCGCGCCAGCGCGCG 36534
Qy 1659 ProThrTyrProHisLeuTyrProTyrLeuLeuArgGlyTyrProAspThrAlaAla 1678
Db 36535 CCAGGATCCCTCAC-----CCAGATCCGCGGTGCGCCCAACACCCACCTTCGCGCA 36588
Qy 1679 LeuGluAspArg-GlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAs 1698
Db 36589 ATCCCGACCGGTGATCCGGTC-----GCCAGGCTTCGATCACCG 36630
Qy 1698 nThrAlaThrAla---MetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgG1 1717
Db 36631 -GGTCCACGCGCACTCCGCGGTACCGTCCGCGCGTCCGCGCGCGGAGCGTTCGCGCA 36689
Qy 1717 uSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnVa 1737
Db 36690 G-----CGACCGCGCGCGCCAGCC----- 36708
Qy 1737 lProHisLeuProValLeuValProThrProGlyThrProAlaThrAlaMetAspAr 1757
Db 36709 -----GTGCGATTCTCCGGCTCTCAGCTCAGAGACGCTCTCGCGCGCG 36755
Qy 1757 gLeuAlaTyrLeuProThrAlaProGlnPro-----Ph 1768
Db 36756 AGGCGCAGTCTTCCCGCGCGCGCCCGCCACCATGTCATGCGAGGTGATCGAGTCTCCG 36815
Qy 1768 eSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysPr 1788
Db 36816 CCGCGTCCGAAGAAGCGCTCGTCCAGCTGACCC---GCTCCAGCGCGCAGCACCTCCGCG 36872
Qy 1788 oThrThrThrSerSerSerGluArg-----GluArgAspArgAs 1801
Db 36873 AACAGCGCGCACAGCACTGTTCCGCTCGCTACCGCTCGCGCGCGCGCTCCCGCGCA 36932
Qy 1801 pArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThrVa 1821
Db 36933 CGCGCATCGGGTCGGGACGCGCTCGCTCGATCTTCCGTTCCGGGTGGTGGGCGAGC 36992
Qy 1821 lGluHisAlaProIleTyrArgProGlyThrGluGlnSerSerSerGlySerSerSe 1841
Db 36993 GCTTCCAGCACCA-----CCAGCGCGGCGCACCATGTGTGCGGCGACCGCGAGCG 37043
Qy 1841 rGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGln----- 1859
Db 37044 AGGTACTCCGTACCGCGGTGCGGATCGACCTCGTTCGGCGACACATGGGCGAGGCGG 37103
Qy 1860 -----HisSerProIleSerProArgThrGlnAspAlaLe 1871
```

Db	37104	TCGTCGGGAGCGTCAACAGCGCTTGGCGGACCTGGCGGTCGGCGACGACGCGACTCG	37163	Db	38136	ACGAACGGCGCTCCGGCCAGCGGACACCCGCCAGCATCCCGACACCCACACGCGCG	38190
Qy	1871	uGlnGlnArgProSerValLeuHisAlaThrGlyMetLysGlyLeuAlaValG1	1891	Qy	2120	uGlnThrAlaPro-----GlyValLysGlyHisGlnArgValThrLeuAlaGlnHi	2138
Db	37164	ATCTCGCGGGCT-----CGATCCGGAACCCAGGAGTTCACCTGGTCAATCG	37211	Db	38191	-----ACCGCCCCACCGCGAGGCCACCGCGG-----	38217
Qy	1891	uProSerLysProThrValLeuArgSerThrSerSerPro-----	1906	Qy	2138	sileSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaPr	2158
Db	37212	GCCGCGCCGCGGAACCACTACCGCTGCTCGTCAGTGGGCGCCAGGTCTCTGTCCGG	37271	Db	38218	-----ACTCCGTCGCGACCCCGCGGCGCTCAGCAGC-----	38249
Qy	1907	-----ValArgProAlaAlaThrPh	1913	Qy	2158	oLeuProLa-----ProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLe	2176
Db	37272	TACATCCGCTCACCGGGGACGAGGATCGCGACGATCGTCCCGCTCAGCGCGCA	37331	Db	38250	GCGGCGAGCGGACGCGCTTCCAGAGCTCCCGAAGGTCACT-----CGCG	38300
Qy	1913	eProProAlaThrHisCysProLeu-----	1921	Qy	2176	uArgArgProProSerAspLeuTyrLeuProProAspHisGlyAlaProAlaArgG1	2196
Db	37332	CGGCCAGGTAGCCACGGGCAACCGGATCCCGACACGATAGAGTTCACCGGTGACCCGC	37391	Db	38301	TCGCGCTCCACCA-----CCGCGTCTCGTGGCGCGCTCTCTGAAGCACC	38345
Qy	1922	-GlyGlyThrLeuAspGlyValTyrProThrLeuMetGlu-----ProValLeuLeuPr	1939	Qy	2196	ySerProHisSerGluGlyLysArgSerProGluProAsnLysThrSerValLeuG1	2216
Db	37392	GGCGGCAACCGCGGAAGAACGGCTCCAGTACGTGGCAGCGGCTTCCGGGATCGGCTCC	37451	Db	38345	CGCGCTCACCGGCGCGCGCCATTCGCGCACCGCCACACCGCGAAGGTTCGGCG	38453
Qy	1939	oLysGluAlaProArg-----	1944	Qy	2232	rGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnTh	2252
Db	37452	CCGATCGGACCACTCCCGCGCGCGTCAACGGCATGCTCATCGCGGCACACACCGTG	37511	Db	38394	CGCGCACCGACCGGCGGACTCCCTCGCGCGCGCGCGCGCGCG-----	38394
Qy	1945	-----ValAlaArgProGluArgProArgAl	1953	Qy	2216	yGlyGlyGluAspGlyLeuProVal-----SerProPro-----GluGlyMetTh	2232
Db	37512	GCCTCGGTCGGCGCTAGCGGTTCATCAGCGCTGCTCCCTCGGACCGGTCGCGGAGG	37571	Db	38395	-GGCGCTCACCGGCGCGCGCCATTCGCGCACCGCCACACCGCGAAGGTTCGGCG	38453
Qy	1953	aAspThrGlyHisAlaPheLeuAlaLysProProAla-----	1966	Qy	2232	rGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnTh	2252
Db	37572	CCCGCGGGGAGGCTCCCTGTCGCCACACCACTCTCCAGGCTCTCCGGCAGCAGCTCG	37631	Db	38454	GCACCGCGCGGCGCGGCGGACCACTCCAGCGTGACTCGG-----CGCGGT	38507
Qy	1966	gSerGlyLeuGluProAlaSerProSerLysGly-----	1981	Qy	2252	rGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePh	2272
Db	37632	GCGTGGCGGACCGACGCGGCGGACCGTCACTGGGTGACGTCGCCACCGCTCCAGTGCC	37691	Db	38508	TCCTCGATCTCGCGTGGCGGTAGCGAGCTGATCTCAGGTGGTCTCCCGCGGATGATG	38567
Qy	1981	oArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAs	2001	Qy	2272	eSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnLeuLeuLys	2292
Db	37692	TCGCCACGACACACCGCGCGCGAGTCTGCTTCGGTG-----	37731	Db	38568	CGACGGTCAGCGGATAGTGGTCCCTCGACCGTCCCGATAGTAGTGTGTCGGCG	38627
Qy	2001	nLeuAlaProHisAlaSerProAsp-----ProProAlaProProAlaSerAl	2018	Qy	2292	sLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrG1	2312
Db	37732	-----CCATCACCGCTGCGCCCGGAGCGCCATCAGCGCTCCGAGACCGTG	37784	Db	38628	AGTCCCGGGGTAGTCTTCGAAAGCAGCAGCGGTCTCGAACTTCGCGCGCGCGCGCG	38687
Qy	2018	aSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnLeuGluLe	2038	Qy	2312	uLlePheAsnMetProAlaIleThrGlyLeuMetThrTyrArgSerGlnAlaVa	2332
Db	37785	GCGTCGAAGCGCGCGGAACTGACGAGGACCGTGCAGCAGCGGACGAGCGAACCGG	37844	Db	38688	ACCTCTGATCTCCGCGAGTCCAGGTCTGATCGGGATCAGGCGCGCTGCGCGCGC	38747
Qy	2038	uArgSerLeu-----GlyTyrHisGlySerSerTyrSerProGluG1	2052	Qy	2332	lGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleAlaLysAlaLeuMetG1	2352
Db	37845	TCGATCTCGCGCGCGCGAGATTCCCGAGCGCGCGTGGGTGACGACGACACCTTCGGG	37904	Db	38748	TCGAGACCGGTTCAGCAGATCGGGA-----CGTCTCGCGCGCTCGAGCGACACCGG	38801
Qy	2052	yValGluProValSerProVal-----SerSerProSe	2063	Qy	2352	yLysTyr-----AspGlnTrpGluGluSerProProLeuSerAlaAsnAl	2367
Db	37905	GTTC-----CCGTGACCGCGAGTATAGATACATACGCCAGCTCGTGGCGCCACCGC	37961	Db	38802	ACCGCACCGGTGTTTCATCAGCCCGGCGATCGACTCCACCCCGGAGTCCCGCGGCG	38861
Qy	2063	rLeuThrHisAspLysGlyLeuProLysHisLeuGluLeuAspLysSerHisLeuG1	2083	Qy	2367	aPheAsn-----ProLeuAsnAlaSerAlaSerLeuProAlaAl	2380
Db	37962	GGCAACTCCGCATCCGGGACCGCGGAG-----GATCGACCTCGTGCATCACCGAGC	38015	Db	38862	CGGACGCGACCGGTGTGCGGAACACCACTCGTGTGCGCGCGCGCGCGCGCGCACCC	38921
Qy	2083	uGlyLeuLeuArgProLysGlnProGlyProVal-----	2094	Qy	2380	aMetProLleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyG1	2400
Db	38016	CGGTCCGCGAATCTCTCGGACCGCGCTCCCGTCCCGTCCCGTCCCGGATCCACCGACG	38075	Db	38922	ATCGCCC-----ACGCGCTCGCACCGCGTGTTC	38951
Qy	2095	-----LysLeuGlyGlyGluAla-----	2102	Qy	2400	yGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaPro--	2419
Db	38076	GGATCCGGCTCGCCAGAGGAGGCTCAGCGCTCGCGCGGATACCGGGATCCACCGGA	38135	Db	38952	AGCGTCAGCCCGGTGTCAGGGCCACTGCGTCAACCGGGTGTGGAGTCCCGGACGACG	39011
Qy	2102	sLeuProHisLeuArgProLeu-----ProGluSerGlnProSerSerProLeuLe	2120	Qy	2420	-----GlyLeuAlaSerGlyAspArgProProSerValSerValHisSerG1	2436
Db	38076	GGATCCGGCTCGCCAGAGGAGGCTCAGCGCTCGCGCGGATACCGGGATCCACCGGA	38135	Db	39012	TAGCGATGCGGTACTCGTATCCCGTCCCGCGCGCTCTTCGCGATCGACGGGTGCTCG	39071
Qy	2102	sLeuProHisLeuArgProLeu-----ProGluSerGlnProSerSerProLeuLe	2120	Qy	2436	u-----GlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpG1	2451
Db	38076	GGATCCGGCTCGCCAGAGGAGGCTCAGCGCTCGCGCGGATACCGGGATCCACCGGA	38135	Db	39072	TCGAGCGCGGACAGCTCCGCTCCAGGACGCGGTGCGCGCTCTGTCTCTCCGACCC	39131

/note="RV0588, (MTCY19H5.34c), len: 295 aa. YrbE2B, hypothetical unknown integral membrane protein, part of mce2 operon and member of YrbE family (see citations below), highly similar to Mycobacterium tuberculosis proteins O07413|RV0168|MTCI28.08|YrbE1B (289 aa); O53966|RV1965|MTV051.03|YrbE3B (271 aa); etc. Also highly similar to conserved hypothetical integral membrane proteins of the yrbE type, e.g. P45392|YRBE_ECOLI hypothetical 27.9 kDa protein from Escherichia coli (260 aa), PASTA scores: Opt: 232, E(1): 8.4e-08, (22.1% identity in 267 aa overlap); P45030|YRBE_HAEN|H11086 hypothetical protein from Haemophilus influenzae (261 aa), PASTA scores: Opt: 234, E(1): 6.3e-08, (24.2% identity in 215 aa overlap); NP 302655.1|NC_002677 conserved membrane protein from Mycobacterium leprae (289 aa); etc."

/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN YRBE2B"
/protein_id="CAB09950.1"
/db_xref="GI:2222770"
/translation="MVESSTASAAVLRARYPRTAASLDYRGVGGTARRLERTGTFAFRTISVQIGHALRRYRRETRLVAEIGMGTGMAVVGTVIIGFTVLSGSLIAIQG FASIGNIGVGAFTGFFAALNTRVAPIVSGVALAATVAGATQALGAMRISSEIDAL EVMGKISFLVSTRILGLVIMPLVALDINAFTSGQVTVTFVFGQSGNRYEHYFR TFLRPEDVGSVVEVVIIVVMITHCYGYGTASGPGVGVQAVGRSFRSLVSVVVV VLLAELALYGVDPNFLTIV"
2548..3762
/gene="mce2A"
/locus_tag="RV0589"
2548..3762
/gene="mce2A"
/locus_tag="RV0589"
/function="UNKNOWN, BUT THOUGHT TO BE INVOLVED IN HOST CELL INVASION"
/standard_name="mce2"
/note="RV0589, (MTCY19H5.33c), len: 404 aa. mce2A; belongs to 24-membered Mycobacterium tuberculosis Mce protein family (see citations below), highly similar to Mycobacterium tuberculosis proteins P72013|MCE1|RV0169|MTCI28.09|mce1A (454 aa); O53967|MCE3|RV1966|MTV051.04|mce3A (425 aa); etc. Also highly similar to others e.g. AAD52105.1|AFI13402.1|AFI13402 mycobacterial cell entry protein from Mycobacterium bovis BCG (454 aa); NP 302656.1|NC_002677 putative cell invasion protein from Mycobacterium leprae (441 aa); CAC12798.1|AL445327 putative secreted protein from Streptomyces coelicolor (418 aa); etc. Also highly similar, but longer 21 aa, to P72013|CAA50257.1|X70901|MTCI28.08 Mcep protein from Mycobacterium tuberculosis (432 aa), PASTA scores: Opt: 1324, E(1): 0, (62.6% identity in 436 aa overlap). Contains a possible N-terminal signal or anchor sequence. Note that previously known as mce2."
/codon_start=1
/transl_table=11
/product="MCE-FAMILY PROTEIN MCE2A"
/protein_id="CAE55301.1"
/db_xref="GI:38490208"
/translation="MPTLVTRKNRRALVYEGVLLLVGLVLLVYKQFGEPTPKT ELTVAFRAGLVMEAGSKVTYNGVEIGRVSGISEIERDGRPAALVLDVNPYISLIP VNVADIEAATFGNKYVALSAPKIPOQORISSHDVIDVGSVTFEFTLFTITISAE KYDPLENTLISAVQALDGLGKFEISITVNGNQLAOLNPLPOLGVDVRLADLGE VYVDASPDWLSFLQNALITARTLTQQRDLDLAALLAATGAGTGEDVFARGGPVLYA AADLVPTATLLDYTSPELFQMRINFDPAKVDVAGNGVSLAAGTILGAPNPVYV PNLPRVNAHGGPGRGPCWCTITRELMPAPVLYVMDTGLASLAPYNHVELGPMFTFYV WGRQYGEINTNP"
3759..4586
/gene="mce2B"
/locus_tag="RV0590"
3759..4586
/gene="mce2B"
/locus_tag="RV0590"

Alignment Scores: 9.34e-05 Length: 349564
Pred. No.: 689.00 Matches: 526
Score: 33.44% Conservative: 207
Percent Similarity: 24.00% Mismatches: 872
Best Local Similarity: 5.21% Indels: 599
Query Match: 1 Gaps: 91
DB:

US-09-522-753-5 (1-2517) x BX842574 (1-349564)

QY 546 LeuLysGluLysThrAspThrSerGlyGluAspAsnAspGluLysGluAlaValAla 565
Db 157203 GTTCGTCGAAGGTGATGGCGGTGACGGCTTCGACGAATGCAATGCTCTGCG 157144
QY 566 SerLysGlyArgLysThrAlaAsnSer-----GlnGlyArg--- 578
Db 157143 TCGCGGTGACGAATCTCGAAGATCGCTCGATTGTCGCCAGCGGGAACAGCGCCA 157084
QY 579 -----LysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSer 592
Db 157083 TACCAGGTTGGCACTGTGAAGGGTCTGACGCCGAGCAGTCGATCGAACCGGGCG 157024
QY 593 Glu-----GluAlaIleThrProGlnGln----- 600
Db 157023 GACGGTTCGGTGATTCGGGTGTCACCGCGGTGTGCAGCCAGCAGCAGCTTGGCGTCCAG 156964
QY 601 -----SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrp 615
Db 156963 AAGGAACATTCGCTGCGCCTATTGGGGGTACTTCCGGGAGTTTTCAGTTCCAGGCC 156904
QY 616 ThrGluGluGluMetGluThrAla----- 623
Db 156903 TCGTCGAGCATTTCGACACACCTCTATTTCGAGGTTCAGTCGATCGCTCGCGCGGACCG 156844
QY 624 -----LysLysGlyLeuLeuGluHisGlyArg-----AsnTrpSerAla 636
Db 156843 GTGCCCGCTCGAATAACCGGACCGGTC-----GGCGCGCGCGCGCGCGCGCGCG 156790
QY 637 Ile----- 637
Db 156789 GCGAATCCCGCGAAGGGCGTCTCGATCAGACGCGCCGCGGATTCACCGCTCGCGG 156730
QY 638 AlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLys 657
Db 156729 GCCCGCGTTCGGCGCGCAGCATTTTCA----- 156700
QY 658 LysArgGlnAsnLeuAspGluIleLeu-----GlnGlnHis 669
Db 156699 -----TCGAGAGATTGACACCGTGGTGGCATGATGCACCGTAAACGATCAACAGCAT 156646
QY 670 LysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysAlaProAlaAla 689
Db 156645 CATCTGGCGTC-CAGCGCGCGAAGCATGTGTCAACGCGCGCGCGCGCGCGCGCG 156587
QY 690 AlaSerGluGluAlaAlaPheProValValGluAspGluMetGluAlaSerGly 709
Db 156586 CGGTATCATCAAGCGTTTCAGCCCGCTTCACCGCAGCAGCGCGCGCGCGCGCGCG 156527
QY 710 ValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsn 729
Db 156526 GCCCGCTGG----- 156518
QY 730 GluValProArgGlyGluCysSer-----GlyProAlaThrValAsnAsnSer 745
Db 156517 -----GCCCGGTGCACCCCGGATTTCCCGCGGTACCCTGTGTACCGCGTT 156470
QY 746 SerAspThrGluSerIleProSerProHisThrGluAlaAla----- 759
Db 156469 GCCCGCGTTGCCCGCTTCGCGATCAGCAGCGCGCTGCCCGCGCGCGCGCGCG 156410
QY 760 -----LysAspThrGlyGlnAsnGly-----ProLysProProAlaThrLeuGlyAl 775
Db 760 -----LysAspThrGlyGlnAsnGly-----ProLysProProAlaThrLeuGlyAl 775

Db 156409 GATGTGAACACCGGACAAACCGGTGGCGCCCGCGCGTCACTCCGAGACC 156350
Qy 775 aAspGlyPro-----ProProGlyProProThrProPro-----ArgArgThrSerAr 791
Db 156349 GAACGCTCCGACATTGCGCGCGCGCGCGTCCCGCGGTGGCGCGCGTGGCGGA 156290
Qy 791 gAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProAl 811
Db 156289 GCGCGCGCGCGCGCGCGCGCG-----CGGAGCGCGACGAAACCGCGTCCGCGCGCG 156233
Qy 811 aProProSerProSerAlaPro-----ProProValProLysGluGluLys 827
Db 156232 CCCCGCGCGCGCGCGCGCGCGTGTGTAACCCCGCGTCCCGCGCGCGCGCGCG 156173
Qy 827 sGluGluGluThrAlaAlaAlaProProValGluGluGluGluGluGluGluGlu 847
Db 156172 GCCAAGACGACGCTCCGCGCGCGCGCGATG-----CGGCC 156137
Qy 847 aAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysTh 867
Db 156136 GACCACCGCGCGCGCGTCAAG-GCTGGAACCGCG----- 156104
Qy 867 rGluGluAlaGluGluGluProAlaLysGlyLysAspAlaGluAlaAlaThrAl 887
Db 156103 -----GCTGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 156051
Qy 887 aGluGlyAlaLeuLysAlaGluLysLysGluGlySerGlyArgAlaThrAlaLys 907
Db 156050 CGGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 155991
Qy 907 sSerSerGlyAlaProGlnAspSerAspSerAlaThrCysSerAlaAspGluValAs 927
Db 155990 TGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 155931
Qy 927 pGluAlaGluGlyClyAspLysAsnArgLeuLeuSerProArgProSerLeuLeuThrPr 947
Db 155930 TGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 155871
Qy 947 oThrGlyAspProArgAlaAlaAlaSerProGlnLysProLeuAspLeuLysGlnLeu 967
Db 155870 CCGCGCGCGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 155814
Qy 967 sGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProProArgGl 987
Db 155813 -CCCGGTG-GCGAACAGCGCGCGTTA-----CGCGCGCGCGCG 155777
Qy 987 uAspAlaAlaProThrLysProAlaProProAlaProProProGlnAsnLeuGlnPr 1007
Db 155776 GCCATTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 155737
Qy 1007 oGluSerAspAlaProGlnProGlyLysSerProArgGlyLysSerArgSerProAl 1027
Db 155736 -----GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 155696
Qy 1027 aProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProPr 1047
Db 155695 CCGCGCATTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAA 155636
Qy 1047 oCysTrpThrSerGlyLeuProPheProValProProArgGluValLleLysAlaSerPr 1067
Db 155635 GAGCAAGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 155576
Qy 1067 oHisAlaProAspProSerAlaPheSerTyAlaProProGlyHisProLeuProLeuGl 1087
Db 155575 GCTGCCCGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 155516
Qy 1087 y-----LeuHisAspThrAlaArgProValLeuLeuProArgProThrIleSerAsnPr 1105
Db 155515 CCGCGCGTTCGAGCGCGCGCGTTCCTCGCGCGCGTTCCTCGCGCGTTCGCGAACAACCA 155456
Qy 1105 oProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIl 1125
Db 155455 GCGCGCGCGTCCGCGCGTCCGCGCGCGTCCGCGAGTGCC-----GTTGCGCGCGT 155406

Qy 1125 eSerGln-----GlyMetSerValGlnLeuHisValProTySerGluHisAl 1141
Db 155405 TGCGGATCAGCGGACGCCCGGTGAGCGTCTGGAGAGGGTCTTACCATATTGATCAGT 155346
Qy 1141 a----- 1141
Db 155345 CCTGCTCAGGGTGTGAGGGGCGAGGTGTGGCGGGGCGGTAAAGCCGCTCCAGGCGCA 155286
Qy 1142 -LysAlaProValGlyProValThrMetGlyLeuLeuProMetAspProLysLysLe 1161
Db 155285 ACAGACCCCGCGTACCAGCGTCCCGCGCTTGGCCAGAGTCCCGCGGTCCCGCGCG 155226
Qy 1161 uAlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProPr 1181
Db 155225 TGCGCGGTTCGCGCGTTCGCGATCAACACCGCGGTGCGCGCGCGCGCGCGCGCAC 155166
Qy 1181 oGluSerLeuGlyValProThrAla-----GlnGluAlaSerValLeuArgGlyThrAlaLe 1200
Db 155165 C-----GTCCCGCGCGGTGTGCGCGCGCGGTTCGCGCGCGTTCGCGCGGT 155118
Qy 1200 uGlySerValProGlyGlySerIleThrLys-----GlyIleProSerThrArgVa 1217
Db 155117 TGCGGATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTTAC 155058
Qy 1217 lProSerAspSerAlaIleThrTyArgGlySerIleThrHisGlyThrProAlaAspVa 1237
Db 155057 TAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 155003
Qy 1237 lLeuTyLysGlyThrIleThrArgIleIleGlyLysAspSerProSerArgLeuAspAr 1257
Db 155002 -----GCGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 154971
Qy 1257 gGlyArgGluAspSerLeuProLysGlyHisValIleTyLeuGlyLys----- 1273
Db 154970 TGCGACTAGCGCGCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCC 154911
Qy 1274 ---LysGlyHisValLeuSerTyArgGlyGlyMetSerValThrGlnCysSerLysGl 1292
Db 154910 TGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCC 154852
Qy 1292 uAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyAs 1312
Db 154851 CTTCCGCGCGCAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 154794
Qy 1312 pMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGl 1332
Db 154793 ---CGCGGTTCGAGAAATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 154738
Qy 1332 yArgAlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHisIleAr 1352
Db 154737 GCGTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAG 154684
Qy 1352 gGlySerIleThrGlnGlyIleProArgSerTyValGluAlaGlnGluAspTyLeuAr 1372
Db 154683 G-----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 154645
Qy 1372 gArgGluAlaLysLeuLysArgGlyThrProProProProProProProProProPro 1392
Db 154644 CCGCGTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 154594
Qy 1392 pLeuThrGluAlaTyThrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGl 1412
Db 154593 -----CGGTG----- 154588
Qy 1412 uGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGl 1432
Db 154587 -----CCGATCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 154562
Qy 1432 uLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGlySerIleTh 1452
Db 154561 GTTGGCTCCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 154519

QY 1452 xGlnGlyThrProLeu-----LysTyrAspTh 1461
 Db 154518 -ACGGGGCGCCGTCGCCGCCACGAATTCGGCTTGATCGGGCCAGCGGCGACAC 154460
 QY 1461 rGlyAlaSerThrThrGlySer-LysLysHisAspValArgSerLeuIleGlySerProG 1481
 Db 154459 GCGCGGCGCTCGCGCGCGCATACGCGCTCGCGCGCGCTCAGGGCTGCACGAACGTG 154400
 QY 1481 lyArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluA 1501
 Db 154399 GGGCTAAACCGCGCGG-----CCTGTG----- 154377
 QY 1501 xGalaCysTyrGluGluSerLeuLysSerArgProGly-----ThrAlaSerS 1517
 Db 154376 -----CGTCCGCCCTGGTATGGCTTGCTGGTGGCCACCGAACCGCGC 154331
 QY 1517 erSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProA 1537
 Db 154330 AATCGCTGTGCACACCTCATCGCACCGCGCGGTCAACAATCCGTCGTGTGACGCGCG- 154272
 QY 1537 rgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuPro----- 1555
 Db 154271 -----CGCGCGCATATAGCGAGCT 154253
 QY 1556 -----ArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluG 1572
 Db 154252 AATCGCGGACCGCATCCCGGCCAAATCCGCTGCGCGCGCCACACCATCTCCGCGGAAC 154193
 QY 1572 lySerLeuSerSerLysAlaSerGlnAspArgLysLeuThrSer----- 1587
 Db 154192 CATCACCATGACATCGATCGCTCCCAA-----CCCTCATCGCGCTGTGCGCGACTA 154139
 QY 1588 -----ThrProArgGluIleAla-----LysS 1595
 Db 154138 CCAGCGCACACCGGTACGGGTAGGAATTAATTCAGAAATCAACGATATCGCGATCCGG 154079
 QY 1595 exProHisSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuL 1615
 Db 154078 CACCGCATGACGGCTTCCACACGTTCGCAG-----GGCCAGCTGCTGGACGGC 154028
 QY 1615 euArgGlyValSerGlyValAspLeuTyrArgSerHisIlePro-----LeuAlaP 1632
 Db 154027 TCACGGCTCGCGCGCGGTTCGCCACAAACGGCGCGGTGCGCGGCACAGGCGCGCGC 153968
 QY 1632 heAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuP 1652
 Db 153967 CGAAACCCGCA-----CCGCGCGGATAGCCGTCTCAGCAGACCGCGCG----- 153924
 QY 1652 roArgHisLeuAlaProAsnProThrTyrProHisLeuTyr-ProProTyrLeuIleArg 1671
 Db 153923 --ACGGGCTCTGACAAGCGCTCAGCGCGCGGGATCTTCGCGAGATACCTGGCGCG 153866
 QY 1672 GlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThr 1691
 Db 153865 CG-GCCGCCATTTGGCGGCGCG-----TCC 153840
 QY 1692 SerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArg 1711
 Db 153839 GCAGCGCATCTGCATCACCGCGGAGTGGTGACGCCCGAGCCATCCGTCGCGCGCGCG 153780
 QY 1712 GlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGly-Pro-----Ar 1729
 Db 153779 CTACGGCAACCGTTGAGCCGCTTTGGCCAGCAGCAACCGCGCGCGCGCGGTTC 153720
 QY 1729 gGlyIleAlaAspLeuSerGlnValProHisLeuProValLeuValProProThrProGl 1749
 Db 153719 GGGGCTCGCGCGTGTGGCGCGGTGCGCGCTTCCCGCGGTG-----CCGCGGTGCGCAT 153663
 QY 1749 yThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSe 1769
 Db 153662 CAGCGCGCGCTGCGG-----CCGACGCCACCGCGCGCG----- 153629
 QY 1769 rSerArgHisSerSerProLeuSerProGlyProThr-His-----LeuT 1786

Db 153628 -CCGAAGCCAGCCAGCGCGCATTCGCGCGGCCCCACCGACCCAGCGCGGTGCCGA 153570
 QY 1786 hrLysProThr-----ThrThrSerSerGluArgGluArgAspA 1800
 Db 153569 CCGCGCAACCGCGCGCGGCCCGCCGCTGGACAAACAGCAGCCGCGGTTCGCCACCGCGCC 153510
 QY 1800 rgAspArgGluArgAsp-----ArgAspArgGluArgGluLysSerIleLeuThrS 1817
 Db 153509 GCGTCCCGCACGAGAAACAAATCCGCGCGCGCGCGCGCGCGCGCGCTGCCGACAG 153450
 QY 1817 erThrThrThrValGlu-----HisAlaProIleTyrArgProGlyThrGluGln-Ser 1834
 Db 153449 CAGCGCTGGCTACCGCGCGGCCCGCGCTCCACCGGTGTCCAGGTTTCGCGGTCCACC 153390
 QY 1835 SerGlySerSerGlySerSer-----GlyGlyGlyGlyGlySerSerSerArgProAla 1852
 Db 153389 GCGGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCATTCGCCCGCGCGCGCGCGCC 153330
 QY 1853 SerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGln 1872
 Db 153329 AACCCTCGCTGCCATTAGTCTGAGCCCGCGGTCCCGCGGTGACCTCCGCGCGCAC 153270
 QY 1873 GlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluPro 1892
 Db 153269 GATCAGCCCGCG-----CAGCAGTCCGCGGTCCCGCGCGC-----CCG 153231
 QY 1893 SerLysProThrValLeuArgSerThrSerThrSerSer-ProValArgProAlaAlaTh 1912
 Db 153230 ACCGCGCGCGCGCTC-----GCCACCAACAGCGCGGTTCACCGCGCGCGCGCC 153183
 QY 1912 rPheProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLe 1932
 Db 153182 GGGCCACCGCGCGCGCCACAGCCCA-----CCGACCCC 153150
 QY 1932 uMetGluProValLeuLeuProLysGluAlaProArgValAlaAlaArgProGluArgProAr 1952
 Db 153149 GCCTCCCGCGCGGTCCGCGCGAGT-----ACGGTGTGGCTCCGCGCATCCCGCC 153099
 QY 1952 galaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAl 1972
 Db 153098 GGGCCCGCGCTGCGCGAGC-----AGCAGCGCGCGCGCGCGCGCGCGCGC-----CCACCGCG 153048
 QY 1972 aserSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAl 1992
 Db 153047 CCGCGCGCACCGCGCGAGTCTCCGCGCGCGCGCGCGCGCGCGCGGTATCC-----CACAG 152991
 QY 1992 aThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProPr 2012
 Db 152990 CCGG---GCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152934
 QY 2012 oAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSe 2032
 Db 152933 GGGCCCGCGTCCCGCAGCAGCCACCA----- 152906
 QY 2032 rIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGl 2052
 Db 152906 ----- 152906
 QY 2052 yValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLy 2072
 Db 152905 -----CCGGGGCGCGCGGTGCC-----CCGCTGCCGACCGCGCGCGCGGTGCCG-- 152861
 QY 2072 sHisLeu-GluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProG 2092
 Db 152860 -----TTGCCGATCAGCGTCCCGCGCGGTGCTGGTGGCGCGCTGTATCGCGCGCAG 152805
 QY 2092 lyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluS 2112
 Db 152804 CGCTGT---GTTTCAGGGTGTGACGCGAGAG-GCGCTGGCGCGGGGTGTTGAAGCGGT 152749
 QY 2112 erGlnProSerSerProLeuGlnThrAlaProGlyValLysGlyHisGlnArgV 2132


```
Db 152748 CGCGCCAGCAGAGCCCACTGATCCACCCGCGCG----- 152711
Qy 2132 alValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisP 2152
Db 152710 -----GTATCACCGCGTTCGTCGGT-----CGCG 152683
Qy 2152 roGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysP 2172
Db 152682 CGATCGCGCGTTCGCGCGTTCGCGATCAGCACCCCGTTCGCGCGCGCGCGCGGTAC 152623
Qy 2172 roValLeuAspLeuArgProPro-----SerAspL 2183
Db 152622 CGCGCGTACAGCGCGCTTGTCCACCGCGCGCGCGCGCGCGCGCTTACCGATCACCAACC 152563
Qy 2183 euTyrLeuProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyG 2203
Db 152562 CGGCTTGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152503
Qy 2203 lyLysArgSerProGluProAsnLysThrSerVal-----LeuGlyGlyGlyGluA 2220
Db 152502 CGCGGACCCACCGGAGCCCAACACAGCC-GGCGTACCGCGCGCGCGCGCGCGCGTCCCGCG 152444
Qy 2220 spGlyIle---GluProValSerProProGluGlyMetThrGluProGlyHisSerArgS 2239
Db 152443 TCGGTATCGCGGAACCCCGCGTCCCGCGGATCCGCGGTGCGGAACAGACACACCGGTG 152384
Qy 2239 exAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSer----- 2255
Db 152383 TTGCGCGCGCGCGCG-----ACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152348
Qy 2256 -----ArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheS 2273
Db 152347 AGCGCGGCTCGCGCGGATCCACCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152288
Qy 2273 erLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLysL 2293
Db 152287 CCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152258
Qy 2293 ysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGly----- 2310
Db 152257 AAGCCAGTCCGCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152198
Qy 2311 --ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerG 2330
Db 152197 CCACCGGCTCGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152139
Qy 2330 lnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaL 2350
Db 152138 GCGGTGCCCGCGCGCG----- 152122
Qy 2350 euMetGlyLysTyrAspGlnTyrGluGluSerProProLeuSer-AlaAsnAlaPheAsn 2369
Db 152121 -----CCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152097
Qy 2370 ProLeuAsnAlaSerAlaSerLeuPro-----Ala 2379
Db 152096 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152037
Qy 2380 AlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeu----- 2394
Db 152036 CAACAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 151977
Qy 2395 -----ThrSerProGlyGly-----GlyGlyLysAlaLysValSer 2406
Db 151976 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 151917
Qy 2407 -GlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspAr 2426
Db 151916 GGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 151860
Qy 2426 gProProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuTh 2446
Db 151859 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 151815
```

```
Qy 2446 rAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPhe---ProTyrAs 2465
Db 151814 GCGTTCGCGATCAACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 151755
Qy 2465 nProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuPr 2485
Db 151754 CAGTCTCTGTCACGCTCTCGCGCGGTTCGCGCGCGGTTCGCGCGCGCGCGCGCG 151696
Qy 2485 oAlaGlySerGlyPro 2490
Db 151695 CGTGGCGTAGGCCCC 151680

RESULT 76
E38020 E38020 E38020 30690 bp DNA linear PAT 31-JAN-2002
LOCUS Avermectin aglycon synthase gene.
DEFINITION E38020
ACCESSION E38020
VERSION E38020.1 GI:18626909
KEYWORDS JP 200245457-A/1.
SOURCE Streptomyces avermitilis
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 30690)
AUTHORS Omura, S. and Ikeda, H.
TITLE Avermectin aglycon synthase gene
JOURNAL Patent: JP 200245457-A 1 12-SEP-2000;
THE KITASATO INSTITUTE
COMMENT OS Streptomyces avermitilis
PN JP 200245457-A/1
PD 12-SEP-2000
PF 24-FEB-1999 JP 1999046961
PR
PI SATOSHI OMURA, HARUO IKEDA
PC C12N15/00.A61K31/70.C12N1/15.C12N9/88.C12P19/62, PC
C12Q1/68//C07D493/22,
PC (C12N1/15.C12R1.465), (C12N9/88.C12R1.465), C12N15/00 CC
FH Key Location/Qualifiers
FT CDS (1)-(11916)
FT CDS (11971)-(30687).
FEATURES
source 1..30690
/organism="Streptomyces avermitilis"
/mol_type="genomic DNA"
/db_xref="taxon:33903"
ORIGIN
Alignment Scores:
Pred. No.: 1-59e-05 Length: 30690
Score: 685.00 Matches: 704
Percent Similarity: 30.18% Conservative: 300
Best Local Similarity: 21.16% Mismatches: 1193
Query Match: 5.18% Indels: 1144
DB: 6 Gaps: 148
US-09-522-753-5 (1-2517) x E38020 (1-30690)
Qy 44 HisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnArg 63
Db 11193 CATCATCCGTTCCGGTCTGCATCCATCCATGCCAGCCCGGA-----CGC 11234
Qy 64 ArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHis 83
Db 11235 ACTGGCCCTTTCGATGCGGC-----CTGGCTCTGGACCGCGCGTCTGTGTGCC 11285
Qy 84 LeuArgProGluSerHisSerTyr-LeuProGluLeuGlyLysSerGluMetGluPheIl 103
Db 11286 CGCGGACCTCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11336
Qy 103 eGluSerLysArgProArgLeuGluLeuLeu-ProAspProLeuLeuArg---ProSerP 122
Db 11337 GCCCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11396
```


Db 13363 CCGCCACACATCCCGTTCAGACACACCCCGCCGAGCATGTCGCGGAGATCAGCGCC 13422
Qy 685 ys-AlaProAlaAlaSerGluGluAlaAlaA-----PheProPro 698
Db 13423 GACGAGATCCCGTAGTGGCTAGAGCTCTCCGCGGAGTCCAGGGGTGTGCGCGTG 13482
Qy 699 ValValGluAspGluGluMetGluAlaSerGlyValSerGlyAenGluGluMetVal 718
Db 13483 CTGGTGTGGCAAGTCGACCGGCC-----CTGCGC 13515
Qy 719 GluGluAlaGluAlaLeuHisAla-SerGlyAsnGluValPro-----ArgGlyGluCy 736
Db 13516 GCCGAGGCCAGGCCCTGACGCGCCACCTCACCGACACCCCGCGCTCGACTCGCGGAC 13575
Qy 736 sSerGly-----ProAlaThrValAsnAsnSerSerAspThrGluSerI 751
Db 13576 GTCCGGTACACCTCGCCGCGCCGCGCGCTGTCGACACCCCGCCACCTCATCGCC 13635
Qy 751 ePro-----SerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysPr 769
Db 13636 GCCGACCGGACACCTCTCTGCAAGCATCTCCAGGCACCTCGCGGAGGACCCACCCC 13695
Qy 769 oProAlaThrLeuGlyAlaAep-----GlyProProProGlyProPro----- 783
Db 13696 GCCGTATCATCAGCAGCGCCCGCCGCGGAGCGGACCGCGGAGCGCGCAGGAAGACC 13755
Qy 784 -----ThrProProArgThrSerArgAlaProIleGluProThr----- 797
Db 13756 GCATTCTCTCTCCGAGAGGACCGCCCGCCCGCGATGCGCCACCGGCTTACCAAC 13815
Qy 798 -----ProAlaSerGluAlaThrGlyAlaPro----- 806
Db 13816 ACCACCCCGTCTTCGCGCGCGCTCAACGACATCTGCACCCACCTCGACCCCGCCCTC 13875
Qy 807 -ThrProProAlaProProSerProSerAlaProProValProValProLysGluI 826
Db 13876 GACCACCCCTCTCTCCGCTCTCACCAGGACCCCAACACCCAGGACACCCACCCCTC 13935
Qy 826 uLysGluGluGluThrAlaAlaAlaProProValGluGluGluGluGluGluGluProPr 846
Db 13936 GAAGAGCGCGCGCTCTCCAGCAGACCGCGCTACGCGCGCGCGCTCTTCGCGCTTC 13995
Qy 846 oAlaAlaGluGluLeuAlaValAsp---ThrGlyLysAlaGluGluProValLysSerGl 865
Db 13996 CAGTTCGCTCCACCGCTCTCACCAGCGCTACACATCACCACCCCGCTACTACGCC 14055
Qy 865 uCysThrGluGluAlaGluGluGluProAlaLysGlyLysAspAlaGluAlaGluAl 885
Db 14056 GGACACTCTCTCGCGGAAATCAGCGCGCCACCTCGCGGCATCTCACCCTCAGCGAC 14115
Qy 885 aThrAlaGluGlyAlaLeuLysAlaGluLysLys----- 896
Db 14116 GCCACACCTCTACCCCAACGCGCCACCTCTATGAAACATGCCCCCGGACCATG 14175
Qy 897 -GluGlyLysSerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAs 916
Db 14176 ACCACCTCCACACACCCCGCCACCATCACCACCTCACCAGCGCCAGGAAACGAC 14235
Qy 916 pSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnAr 936
Db 14236 CTCGCCATCGCGGCATCAACACCCCGCCACCTCTCTCTCATCAGCGCA----- 14284
Qy 936 gLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAlaAlaSe 956
Db 14285 -----CCCCCACAACGTCACACATCACCA-----CCCTCTGCAACAA 14328
Qy 956 rProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaAlaPro---Pr 975
Db 14329 GGCATCAAAACCAAAACCTCCCAACCAACACCGCTTCCACTCCCGCCACACACCCC 14388
Qy 975 olleGlnValThrLysValHisGluPro---ProArgGluAspAlaAlaProThrLysPr 994

14389 ATCTCAACCACTCCACAGCACACCAAAACCTCTACCTACCAACCCACCCACACCC 14448
Qy 994 oAlaProAlaProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGl 1014
Db 14449 CTATATACCGCAACACCCCGCAACTCTCTACCCCGCTACTGACCAACAA 14508
Qy 1014 nProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAl 1034
Db 14509 GCCGCAACA-----CGTCTGATACGCAACCAACCAACCAACCTCCACCA 14556
Qy 1034 aPheAlaAlaGluAlaGlnLys-----LeuProGlyAspPro---ProCy 1048
Db 14557 CACGCGCTCCACCTATATCGAACTCGAACCCCGCAACCAACCTCCACACCTCCAC 14616
Qy 1048 sTyrThrSerGlyLeuProPheProValPro-----ProArgGluValI 1063
Db 14617 GACACCTCTCCCAACACCCCGCAACCTCTACCTCACCACCCCGCCACCAACCC 14676
Qy 1063 eLysAlaSer-ProHisAlaProAspProSerAlaPheSerTyrAlaProProGly---H 1082
Db 14677 CAAACCCACCTCTCTCAACACCTCGCAAAACCAACCAACCTCTGCGACCCCGCCAC 14736
Qy 1082 isProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrI 1102
Db 14737 ACCACACCA-----CAACCAACCCACCAACCAACCTCTGACCTCCCGCC 14787
Qy 1102 leSerAsnPro---ProProLeuLysSer-----SerAlaLysHisProSerV 1117
Db 14788 TACCCCTTCAACACCAACCTCTCTGCTCAACCAACCCCGCAACCGCGACCC---G 14844
Qy 1117 alLeuGluArgGln----- 1122
Db 14845 TCACCGAGCGAGCGCGTGAAGCCAGCCACCCATCAACCCCGCTGCGTGTCTC 14904
Qy 1122 leGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGlu-HisAla 1141
Db 14905 GTGGGCAAGTCTCGCAGGAGCGAGACAGAGAGCTGTTCGCGCTGTGCGACCATGCG 14964
Qy 1142 LysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeu 1161
Db 14965 GCGCTGTGTGGGCGATGCCATCCCGAAGTGTTCGGAACAGGCGCTTCAAGAG 15024
Qy 1162 AlaProPheSerGlyValLysGlnGlnLeuSerProArgGlyGlnAlaGlyPro--- 1180
Db 15025 CTGGGTTTGTATCTCTCGCGCAATTACGCTTGTAAATCGACTGTCTGCTGACGTTGAC 15084
Qy 1181 -----ProGluSerLeu-----GlyValProThrAlaGlnGlu---AlaSerVal 1194
Db 15085 CTGCGCTTCCGCGCACGCTGATTTCCGATTTACCCCACTCCGATGCGGCTTTGCCAGTTC 15144
Qy 1195 LeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSer 1214
Db 15145 CTGCGGCGGCGATGTCGCGAGCGGACAGAGC-----AGGACACTGCTCTCCCGCTA 15198
Qy 1215 ThrArgValProSerAspSerAlaIleThr-----TyrArgGly 1227
Db 15199 ACTGCGGTCCCGCGCAGCGCGATCGCCATCGTCGCGCATGCGCTGTGCTGACCCCGT 15258
Qy 1228 SerIleThr----- 1230
Db 15259 GATGTACGAGCGGTGATGATCTCTGCGAGGTGTCAGTGTGGCCATGACGCGATCGGC 15318
Qy 1230 ----- 1230
Db 15319 GGATTCGAGCAACCGTGGTGGGACCTCGACACGCTGTACACCCCGGACCCGAGCAC 15378
Qy 1231 HisGlyThr----- 1233
Db 15379 CACGGAACAGCTACACCCCGAGCGGGATTCCTTTACGACGCGAGCAATTTTCGATCCC 15438
Qy 1234 -----ProAlaAspValLeu----- 1238
Db 15439 GACTTCTTGGTATCAGTCCGCGTGAGGCATCGCGATGGACCGCAGCGCGCTGCTG 15498

QY	1239	-----TyrIysGly	1241	QY	1436	Thr-----ProGluLeuProLeu-----	1441
Db	15499	CTGGAACACGGTGGAGATCGAAGACAGCGCTGCATCAACCCGACAGCCTCGGTGC	15558	Db	16500	ACCGCCCAACATCCCGTCAGACACACCGCCGACGAGCCCGGGAGAGACGCGC	16559
QY	1242	ThrIleThrArgIleIleGly-----GluAspSerProSerArg	1254	QY	1442	-----AlaProArgProLeuLysGluGlySerIleThr	1452
Db	15559	ACACCAACCGCGCTCTTCGCGGCTGACCTACACGACTACCGCGCGCTTCCCA	15618	Db	16560	CGACGATGTTCCGGGGAAGCGCGGACGACGCGCGGTACCGCGGGGAGC-----	16613
QY	1255	-----LeuAspArgGluAspSerLeuProLysGly	1266	QY	1453	GlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLeuLysHisAsp	1472
Db	15619	GCTCCGGCAGGGTTCGAGGGGTATCTCGGCACGGAAGCCGACGATATCGCTCGGT	15678	Db	16614	CACGTGCTCTGCTCGCGCAGTCCAGGGGTGTGGCTGGTGTGGCAGAGTCGA	16673
QY	1267	HisValIleTyrGlu-----GlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSer	1285	QY	1473	ValArgSerLeuIleGlySerProGlyArgThrPheProPro-----	1486
Db	15679	CGTGTGCCTACGCTCTCGGCTCGGAGGTCCGGCCCTCACATCGAC-----	15726	Db	16674	GCGGCCCTGCGCGCCAGGCGCCAGGCTGCACGCCACCTCACCGACCCCGGCT	16733
QY	1286	ValThrGlnCysSerLysGluAspGlyArgSerSerGlyProProHisGluThrAla	1305	QY	1487	-----ValHisProLeuAspValMetAlaAspAlaArgAlaLeuGlu	1500
Db	15727	---ACTGCCTGCTCTCTGCTCCCT--GCTCGCTCTGCA-----CCTGCCGTGTGAGCGCT	15776	Db	16734	CGACCTCGCGGACGTGCGGTACACCT-----CGCCCAACGCGCG	16772
QY	1306	AlaProLysArgThrTyr-----AspMetMetGluGly	1316	QY	1501	ArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSer-----	1518
Db	15777	GCGTTCGGGAGGTTCATGCTCGGCTCGGGTGGGTACGCGTGATGTCAACCCCGGC	15836	Db	16773	CGCG--TGTTGCACACCGCGCCACCTCATCGCGCGACCGCGACACCTTCTTGCAG	16831
QY	1317	ArgValGlyArgAlaIleSerSerAla-----SerIleGluGlyLeu	1330	QY	1519	-----GlyGlySerIleAlaArgGlyAlaProVal	1528
Db	15837	CGGGTTCGTGAGTTCGCGGACGCGGGCTGGCGGTGGACGGCGGTGCAAGCGGT	15896	Db	16832	CACCTCCAGCAGCTCGCGGACGAGAACCCACCCCGCGCTCATCCACGACGCGCCAG	16891
QY	1331	MetGly-----	1332	QY	1529	IleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAla	1548
Db	15897	CTCGCAGCGCTGACGCGACCGCTCGGGTGGGTGCGAATGCTGCTGGTGGAGCG	15956	Db	16892	GCGGACCGCGGACCGGGAGCGCGAGAAAGACCGCATTCATCTGCTCGGACAGGGCA	16951
QY	1332	-----	1332	QY	1549	ProPheAlaGlyHisLeuProArg-----	1556
Db	15957	GCTGTGCGACGCGCGGCTCGGTACCGAATCTCGCGTGTGTGGTGGCAGTGGGT	16016	Db	16952	CCCAACGCGCGCATGGCGCCACGCGCTTACCACACCCCGCTTTCGCGCGCGCAC	17011
QY	1333	-----ArgAlaIlePro-----	1336	QY	1557	-----GlySerProValThr	1561
Db	16017	CAATCAGCAGCGTGCAGACAAACGGGCTGACGCGCCCAACGGCGCTGCCAGAGCGGT	16076	Db	17012	TCAACGACATCTGACCCACCTCGACCCCGCACTCGACCCCGCTTCCCTCCCTCTCTCA	17071
QY	1337	-----ProGluArgHis-----Ser	1341	QY	1562	MetArgGluProThrProArgLeu-----	1569
Db	16077	CATCGCTGCCCTGGCAACGCGGACCTACCCCGCGACGCTCGATGCGGTGGAGGC	16136	Db	17072	CCAGGACCCCAACACCCAGGACACCAACCCCTCGAAGAGGGCGGCTGCTCCAGC	17131
QY	1342	ProHisHisLeuLysGluGlnHisIle---ArgGlySerIleThrGlnGlyIlePro	1360	QY	1570	-----GlnGluGlySerLeuSerSerLysLysAlaSerGln	1581
Db	16137	CCACGGCAC-----CGGCACCATTTTGGGGGACCCGATCGAGGCCCGCTCTCT	16187	Db	17132	AGACCGGTAGCCCGACCGCGCTCTTCGCTTCCAGGTGCGCTTCCACCGCTCTCTCA	17191
QY	1361	ArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLysArg	1380	QY	1582	AspArgLysLeuThrSerThrProArgGluIle-----AlaLysSerPro	1596
Db	16188	CGCCACCTA-----CGG	16199	Db	17192	CCGACGGCTACCATCATCCCCCACTACTACGCGGACACTCTCCGCGCAANTCACC	17251
QY	1381	GluGlyThrProProProProSerArgAspLeuThrGluAlaTyrLysThrGln	1400	QY	1597	-----HisSerThrValPro-GluHisHisProHisProIleSerPr	1610
Db	16200	ACAGACCGCCCGCAACGACCGCTGTGGCTCGATGAAGTCGAACATCGGCCA	16259	Db	17252	CCGCCACCTCGCGGATCTCTCCCTTACCGACCGCCACCTCATCACCAACGCG	17311
QY	1401	AlaLeuGlyProLeuLysLeuLysProAlaHisGlu-----	1412	QY	1610	oTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLe	1630
Db	16260	CGCGCAGGTGCGCGAGGTGTGGGGGTGATCAAGATGTGTGCGCTCGCGAATGG	16319	Db	17312	CCACCTCATGCAACCATCGCCCGGACCATGACCACTCCACCTCCACACCCCGCAC	17371
QY	1413	-----GlyLeuVal-----	1415	QY	1630	uAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaIleTyr--	1649
Db	16320	TCTGTGCGGAGCGTTGATGTGTGATGAGCGCTGCGCGCATGTGGACTGTCGCGG	16379	Db	17372	ACATCACCCACCTCATCGCGCCCGAAGAACGACTCGCATCGCGGCCATCACACC	17431
QY	1416	-----AlaThrValLysGluAla	1421	QY	1650	-TyrLeuProArgHisLeuAla-----	1656
Db	16380	GCGGTGACGTGCTGACGAGACGCTGCTGCGCGCGCGGAGGGCGGCTGCGCGC	16439	Db	17432	CCACCTCTCTGTCATCAGCGGCGACCCCGCCACCGTCCAAACATCACCCCTCTGCC	17491
QY	1422	GlyArgSerIle-----HisGluIleProArgGluLeuArgHis	1435	QY	1656	-----	1656
Db	16440	GCGAGAGTGTATCTGTCGGCGTCAGCGGCGACCAACGCCACGCTCATCTCTCGAAGAGC	16499	Db	17492	AACAACAAGGATCAAAACCAAAACCTCCCGCCCAAAAGCGCTTCACTCCCGCCACA	17551
				QY	1657	-----ProAsnProThr-----TyrProHisLeu-TyrProPro	1668

ORGANISM	Streptomyces avermitilis																	
REFERENCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomyces; Streptomyceae; Streptomycetaceae; Streptomyces.																	
AUTHORS	1 (bases 1 to 30690) Endo,H., Yamaguchi,H., Kanda,Y., Hashimoto,S., Omura,S. and Ikeda,H.																	
TITLE	A method for procuring avermectin derivatives																	
JOURNAL	Patent: WO 0162939-A 1 30-AUG-2001; KIYOMA HAKKO KOGYO CO LTD,THE KITASATO INSTITUTE,HIROFUMI ENDO, HIROYUKI YAMAGUCHI,YUTAKA KANDA,SHINICHI HASHIMOTO,SATOSHI OMURA, HARUO IKEDA																	
COMMENT	OS Streptomyces avermitilis PN WO 0162939-A/1 PD 30-AUG-2001 PF 23-FEB-2001 WO 2001JP001381 PI 24-FEB-2000 JP OOP 47405 PI HIROFUMI ENDO,HIROYUKI YAMAGUCHI,YUTAKA KANDA,SHINICHI PI HASHIMOTO, PI SATOSHI OMURA,HARUO IKEDA PC C12N15/04,C12N15/53,C12N9/10,C12N1/21,C07C327/30, PC C12P17/08 CC A method for procuring avermectin derivatives FH Key																	
FEATURES	source	Location/Qualifiers	Location/Qualifiers	FT CDS	FT CDS	FT CDS	FT CDS	FT CDS	FT CDS	FT CDS	FT CDS	FT CDS	FT CDS	FT CDS	FT CDS	FT CDS	FT CDS	FT CDS
ORIGIN	1..30690 /organism="Streptomyces avermitilis" /mol_type="genomic DNA" /db_xref="taxon:33903"																	
Alignment Scores:	Pred. No.:	1.59e-05	Length:	30690	Score:	685.00	Matches:	704	Percent Similarity:	30.18%	Conservative:	300	Best Local Similarity:	21.16%	Mismatches:	1193	Query Match:	5.18%
US-09-522-753-5 (1-2517) x BD097649 (1-30690)																		
Qy	44	HisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleleGlnProGlnArg	63															
Db	11193	CATCATCCGTTCCGGTCTGCATCCCATGTCGACCCCGGA	11234															
Qy	64	ArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHis	83															
Db	11235	ACTGGCCCTCTTCGATGCGGC	11285															
Qy	84	LeuArgProGluSerHisSerTyr-LeuProGluLeuGlyLysSerGluMetGluPhe11	103															
Db	11286	CGCGGACCTGGTCCCGCCCGCCCGCTCGCCCTG	11336															
Qy	103	eGluSerLysArgProArgLeuLeuLeu-Leu-ProAspProLeuLeuArg---ProSerP	122															
Db	11337	GCCCGCACCGCGCGCGGACACCGCGCACCTACCGTGGTGGCGGACACCGCGC	11396															
Qy	122	roLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeu	142															
Db	11397	CCAGCTGACCGCGCTGGCGCGGAGACACAGAACACACACACCTCTCTCGC	11456															
Qy	142	hrGlyLysLeuGluProValSerProProSerProHisThrAspProGluLeuGlu	162															
Db	11457	CCTGGTCC---GCTCCACATCGCCACCGTCTG-GGCCACACACACCC---GACA	11506															
Qy	162	euValProProArgLeuSerLysGluLeuLeuLeuGlnAsnMetAspArgValAspArg	182															
Db	11507	CCATCCCGCC---GACCGCGCTTCGCG	11533															
Qy	182	luLeThrMetValGlnGlnGlnLeuSerLysLeuLysLysLysGlnGlnGlnLeu-Glu	201															
Db	11534	ACCTCGGCTTCGACTCCCTCACCGCGCTCGAATACGCAACCGGCTCTCCCGCACCCG	11593															

Streptomyces avermitilis	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomyceinae; Streptomycetaceae; Streptomyces.	
1 (bases 1 to 30690)	Endo, H., Yamaguchi, H., Kanda, Y., Hashimoto, S., Omura, S. and Ikeda, H.
A method for procuring avermectin derivatives	
Patent: WO 0162939-A 1 30-AUG-2001;	
KYOWA HAKKO KOGYO CO LTD, THE KITASATO INSTITUTE, HIROFUMI ENDO,	
HIROYUKI YAMAGUCHI, YUTAKA KANDA, SHINICHI HASHIMOTO, SATOSHI OMURA,	
HARUO IKEDA	
OS Streptomyces avermitilis	
PN WO 0162939-A/1	
PD 30-AUG-2001	
PP 23-FEB-2001 WO 2001JP001381	
PR 24-FEB-2000 JP OOP 47405	
PI HIROYUKI ENDO, HIROYUKI YAMAGUCHI, YUTAKA KANDA, SHINICHI PI HASHIMOTO,	
PI SATOSHI OMURA, HARUO IKEDA	
PC C12N15/54, C12N15/53, C12N9/10, C12N9/14, C12N1/21, C07C327/30, PC C12P17/08	
CC A method for procuring avermectin derivatives FH Key	
FT CDS (1) . . (11916)	
FT CDS (11971) . . (30687) .	
Location/Qualifiers	
1..30690	Length: 30690
/organism="Streptomyces avermitilis"	Matches: 704
/mol_type="genomic DNA"	Conservative: 300
/db_xref="taxon:33903"	Mismatches: 1193
	Indels: 1144
	Gaps: 148
Alignment Scores:	
Pred. No.:	1.59e-05
Score:	685.00
Percent Similarity:	30.18%
Best Local Similarity:	21.16%
Query Match:	5.18%
DB:	6
US-09-522-753-5 (1-2517) x BD097649 (1-30690)	
Qy 44 HisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleleGlnProGlnArg 63	
Db 11193 CATCATCCGTTCCGGTCTGCATCCCATGTCACCCCGGA-----CGC 11234	
Qy 64 ArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHis 83	
Db 11235 ACTGGCCCTCTTCGATGCGGC-----CCTGGCTCTGGACCGCGCTCTGCTGCC 11285	
Qy 84 LeuArgProGluSerHisSerTyr-LeuProGluLeuGlyLysSerGluMetGluPhe11 103	
Db 11286 CGCGGACCTGGTCCCGCCCGCCCGCTCGCCCTGCGCCCTG-----CTGCAGGACCTCT 11336	
Qy 103 eGluSerLysArgProArgLeuLeuLeu-Leu-ProAspProLeuLeuArg---ProSerP 122	
Db 11337 GCCCGCACCGCGCGCGCGACACCGCGCACCTACCGTGGTGGCGGACACCGCGC 11396	
Qy 122 roLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeu 142	
Db 11397 CCAGCTGACCGCGCGCTGGCGCGGAGACACAGAACACACACACCTCTCTCGC 11456	
Qy 142 hrGlyLysLeuGluProValSerProProSerProHisThrAspProGluLeuGlu 162	
Db 11457 CTGGTCC---GCTCCACATCGCCACCGCTCTG-GGCCACACACACCC---GACA 11506	
Qy 162 euValProProArgLeuSerLysGluLeuLeuLeuGlnAsnMetAspArgValAspArg 182	
Db 11507 CCATCCCGCC---GACCGCGCTTCGCG 11533	
Qy 182 luLeThrMetValGlnGlnGlnLeuSerLysLeuLysLysLysGlnGlnGlnLeu-Glu 201	
Db 11534 ACCTCGGCTTCGACTCCCTCACCGCGCTCGAATACGCAACCGGCTCTCCCGCACCCG 11593	


```
Db 13576 GTGGGTACACCTCGCCAGCGCGCGGTTCGACACCGCGGCCACCTCATCGCC 13635
|||||
Qy 751 ePro-----SerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysPr 769
|||||
Db 13636 GCGGACGCGACACCTTCCTGCAAGCAGCTCCAGGCACTCCGCGCAGCGCAACCCACCC 13695
|||||
Qy 769 oProAlaThrLeuGlyAlaAsp-----GlyProProGlyProPro----- 783
|||||
Db 13696 GCGGTATCCACAGCAGCGCGCCAGCGGAGCGCGGACCGCGGAGCGCGCAGGAAGACC 13755
|||||
Qy 784 -----ThrProProArgArgThrSerArgAlaProLleGluProThr----- 797
|||||
Db 13756 GCATTATCTGCTCCGACAGGCGCACCCCAACGCGCGCGATGCCACCGCCTCTACAC 13815
|||||
Qy 798 -----ProAlaSerGluAlaThrGlyAlaPro----- 806
|||||
Db 13816 ACCACCCGCTTCGCGCGGCACTCAAGCACTCTGCACCCACCTCGACCCCCACCTC 13875
|||||
Qy 807 -ThrProProAlaProProSerProSerAlaProProProValProValProLysGlu 826
|||||
Db 13876 GACCAACCCCTTCCTCCCTCCTCAGGAGCCCAACACCCAGGAGCACCCACCCCTC 13935
|||||
Qy 826 uLysGluGluGluThrAlaAlaAlaProProValGluGluGluGluGluGluGluProPr 846
|||||
Db 13936 GAAGAGCGCGCGCACTGCTCCAGCAGACCGCTACGCCAGCGCGCGCTTCGCGCTTC 13995
|||||
Qy 846 oAlaAlaGluGluLeuAlaValAsp---ThrGlyLysAlaGluGluProValLysSerGl 865
|||||
Db 13996 CAGTGGCCCTCCACCGCTCTCAGCGAGCTACCACTACCCCGCCCTACTACGCC 14055
|||||
Qy 865 uCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaGluAl 885
|||||
Db 14056 GGACACTCTCCGCGGAATCAGCGCGCCACCTCGCGCATCTCCACCCCTCACCGAC 14115
|||||
Qy 885 aThrAlaGluGlyAlaLeuLysAlaGluLysLys----- 896
|||||
Db 14116 GCCACCCCTCATACCCAGCGCGCCACCTCATGAAACCATGCCCCCGGCACCATG 14175
|||||
Qy 897 -GluGlySerGlyArgAlaThrAlaLysSerSerGlyAlaProGlnLysSerAs 916
|||||
Db 14176 ACCACCTCCACACCCCGCCACCATCACCACCACTCACCAGCGCGCGCAAAACGAC 14235
|||||
Qy 916 pSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnAr 936
|||||
Db 14236 CTCGCCATCGCGCCATCAACACCCCGCCCTCTCTGTCACGCGCA----- 14284
|||||
Qy 936 gLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSe 956
|||||
Db 14285 -----CCCCCACCGCTCCACACATCACCA-----CCCTCTGCCAACAA 14328
|||||
Qy 956 rProGlnLysProLeuAspLeuLysGlnLysGlnArgAlaAlaLlePro---Pr 975
|||||
Db 14329 GGCATCAAAACCAAAACCTCCCAACCAACACCGCTTCCTCCCTCCCGCCACCAACCC 14388
|||||
Qy 975 -lleGlnValThrLysValHisGluPro---ProArgGluAspAlaAlaProThrLysPr 994
|||||
Db 14389 ATCTCAACCACTCCACAGCAGACCCCAACCTCCTCAGCCCGCCCTACTGAGCAACCA 14448
|||||
Qy 994 oAlaProProAlaProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGl 1014
|||||
Db 14449 CTCATCAGCGCAACACCCCGCCAGCACTCCTCAGCCCGCCCTACTGAGCAACCA 14508
|||||
Qy 1014 nProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAl 1034
|||||
Db 14509 GCGCGCAACA-----CCGTGCACTAGCGCCACCCACCAACCCCTCCACCA 14556
|||||
Qy 1034 aPheAlaAlaGluAlaGlnLys-----LeuProGlyAspPro---ProCy 1048
|||||
Db 14557 CAGCGGCTCACCACCTACATCGAAGCTCGGACCGGACCAACCTCACCACCTCACCAC 14616
|||||
Qy 1048 strThrSerGlyLeuProPheProValPro-----ProArgGluValIl 1063
|||||
```

```
Qy 1286 ValThrGlnCysSerIysGluAspGlyArgSerSerGlyProHisGluThrAla 1305
Db 15727 --ACTGCCTGCTCTCGTCCCT--GGTCGCTCTGCA-----CCTGSCCTGTCCAGGCGCT 15776
Qy 1306 AlaProIysArgThrTyr-----AspMetMetGluGly 1316
Db 15777 GCGTCGCGGAGTGTTCATGCGCCCTCGCGGTTGGCGTCACGGTGATGTCAACCCCGGC 15836
Qy 1317 ArgValGlyArgAlaIleSerAla-----SerIleGluGlyLeu 1330
Db 15837 CGGTTCTGCGAGTTTTCGCGGACGCGGCGCTGCGCGTGGACGCGCGGTGCAAGCGGTT 15896
Qy 1331 MetGly-----1332
Db 15897 CTCGGCAGCGCTGACGGCACCAGCTGGGTGAGGTGCGAATGCTGCTGTGGAGCG 15956
Qy 1332 -----1332
Db 15957 GCTTCGGACGCGCGCGCTCGGTCTACCGAATCTCGCGTGTGTGGCAGTGCCTG 16016
Qy 1333 -----ArgAlaIlePro-----1336
Db 16017 CAATCAGGACGGTCGAGCAACGGGCTGACGGGCCCAACGGGCGGTGCCAGGAGCGTGT 16076
Qy 1337 -----ProGluArgHis-----Ser 1341
Db 16077 CATCCGCTCGCTGSCCAACGGGACCTGACCCCGCGCGTGCATCGGTGGAGGC 16136
Qy 1342 ProHisIleLeuIysGluGlnHisIle---ArgGlySerIleThrGlnGlyPro 1360
Db 16137 CCACGGCAC-----CGGCACCACTTTGGCGCAGCCGATCGAGGCCCGCTCTCT 16187
Qy 1361 ArgSerTyrValGluAlaGlnGluAspTyrLeuArgGluAlaIysLeuLeuIysArg 1380
Db 16188 GGCACCTA-----CGG 16199
Qy 1381 GluGlyThrProProProProSerArgAspLeuThrGluAlaTyrLysThrGln 1400
Db 16200 ACAGGACCGCCCGCAACGACCGCTGTGGCTCGATGAGTCGAACATCGGCCA 16259
Qy 1401 AlaLeuGlyProLeuIysLeuIysProAlaHisGlu-----1412
Db 16260 CGCGCAGGCTCGCGCAGGTGTGGCGGGGTTCATCAGATGTTGATGGCGCTGCGGAATGG 16319
Qy 1413 -----GlyLeuVal-----1415
Db 16320 TCTGCTGCCGCGAGCTTGCATGTGGATGAGCCGTGCGCGCATGTGGACTGGTCCGCGG 16379
Qy 1416 -----AlaThrValIysGluAla 1421
Db 16380 GCGGTCCAGCTGCTGACGGAGACGGTCCCTGGCCCGGGGAGGGCGGTGCGGCG 16439
Qy 1422 GlyArgSerIle-----HisGluIleProArgGluGluLeuArgHis 1435
Db 16440 GGCAGGAGTGTCTGCTGCGCGTTCAGCGGCACCAACGCCCACTCATCTCTCGAAGAAGC 16499
Qy 1436 Thr---ProGluLeuProLeu-----1441
Db 16500 ACCGCCCAACAATCCCGTCAGACACACCGCGCAGCAGCCCGCGGAGAGGAGCGCCG 16559
Qy 1442 -----AlaProArgProLeuIysGluGlySerIleThr 1452
Db 16560 CGACGATGTTCCGGGGAGACGGCCCGCGCAGCAGCGGTACCGCGGGGAGC-----16613
Qy 1453 GlnGlyThrProLeuIysTyrAspThrGlyAlaSerThrThrGlySerIysIysHisAsp 1472
Db 16614 GACTGGTCTCTGCTCGCGCAGTCCAGGGGTGTGCGCGTGTGTGTCGCGCAAGTCGCA 16673
Qy 1473 ValArgSerLeuIleGlySerProGlyArgThrPheProPro-----1486
Db 16674 GCGGCGCTCTGCGCGCCAGCGCCCGAGGCGCTGCAACCGCCACCTCATCCGACACCCCGGCT 16733
Qy 1487 -----ValHisProLeuAspValMetAlaAspAlaArgAlaLeuGlu 1500
Db 16734 CGACCTCGCGACGCTGGGTACACCT-----CGCCACGCGCG 16772
Qy 1501 ArgAlaCysTyrGluGluSerLeuIysSerArgProGlyThrAlaSerSerSer-----1518
Db 16773 CGCG--TGTTTCGACCAACCGCGCACCTCTATCGCGCGCGGACCGGACACCTTCTGCAAG 16831
Qy 1519 -----GlyGlySerIleAlaArgGlyAlaProVal 1528
Db 15832 CATTCAGGCACTCGCGCAGGCGAACCCACCCCGCGCTATCCACGAGCGGCCCGCAG 16891
Qy 1529 IleValProGluLeuGlyIysProArgGlnSerProLeuThrTyrGluAspHisGlyAla 1548
Db 16892 GCGGACCGGACCGGAGGCGCGAGAAAGACCGCATTCATCTGTCTCGGACAGGCA 16951
Qy 1549 ProPheAlaGlyHisLeuProArg-----1556
Db 16952 CCACACGCGCGCGCATGGCCCGCTTACCAACACCCCGCTTTCGCGCGCGCAC 17011
Qy 1557 -----GlySerProValThr 1561
Db 17012 TCACGACATCTGCACCCACTCGACCCCACTCGACACCCCGCTCTCTCCCTCTCTCA 17071
Qy 1562 MetArgGluProThrProArgLeu-----1569
Db 17072 CCCAGGACCCCAACACCCAGGACACCAACCCCTCGAAGAAGCGCGCGCTGCTCCAGC 17131
Qy 1570 -----GlnGluGlySerLeuSerSerSerIysAlaSerGln 1581
Db 17132 AGACCCGCTACGCGCGCGCTCTTTCGCTTCCAGGTGCGCTCCACCGCTCTCTCA 17191
Qy 1582 AspArgIysLeuThrSerThrProArgGluIle-----AlaIysSerPro 1596
Db 17192 CGACGGCTACCACTACACCCCGCTACTACGCGGACACTCTCTCGGGAATACCG 17251
Qy 1597 -----HisSerThrValPro--GluHisHisProHisProIleSerPr 1610
Db 17252 CGCCCGACCTCGCGCGCATCTCAGCTCAGCAGGCGCACCCCTCATCACCAACGCG 17311
Qy 1610 oTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLe 1630
Db 17312 CCACCTCATCAACCAACCATGCCCGCGGACCATGACCCCTCCACACCCCGCCACCC 17371
Qy 1630 uAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaIaTyr-- 1649
Db 17372 ACATCACCCACCACTCAGCGCCCGCAAGAACGACCTCGCCATCGCGCCATCAACACC 17431
Qy 1650 -TyrLeuProArgHisLeuAla-----1656
Db 17432 CCACCTCCCTCGTCATCAGCGGCGACCCCGCACACCGTCCAAACACATCACCCACTCTGCC 17491
Qy 1656 -----1656
Db 17492 AACACAAGSCATCAAAACCAAAACCTCCCCACCAAAACGGCTTCCACTCCCGCCACA 17551
Qy 1657 -----ProAsnProThr-----TyrProHisLeu--TyrProPro 1668
Db 17552 CCAACCCCTCTCAACCACTCCACGAGCAGACCCCAACCCCTCCTACCTACCCACCC 17611
Qy 1668 yr-----LeuIleArgGlyTyrProAspThrAlaAlaLeu-----1679
Db 17612 ACACCCCTCATCAGCGGCAACACCCCGACCACTCTCTCCCGCCACTACTGGA 17671
Qy 1680 -----GluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHis- 1697
Db 17672 CCACAACAGCCCGCAACACCGTC-----GACTACGCGCACCAACCCCAACCCCTCCACC 17725
Qy 1698 AsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGlu 1717
Db 17726 AACCGGCGTCCACCTACATCGAAGTCCGAC-----CGACAAACA 17767
Qy 1718 SerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnVal 1737
```

Db	17768	CCCTCA-----	17773
Qy	1738	ProHisLeuProValLeuValProThrProGlyThrProAlaThrAlaMetAspArg	1757
Db	17774	CCACCTCACCCACCAACCTCCCAACACACCCACCAACCTCACCTCACCCAC	17833
Qy	1758	LeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerProLeu	1777
Db	17834	CCCAACC-----ACCAACCCCAACCAACCTCTCTCAACCTCGCCAAACACCA	17884
Qy	1778	SerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerGluArgGlu	1797
Db	17885	CCACCTGGACCCCCACC-----ACTACCCACCAACCAACCAACCC-CAACCCAC	17937
Qy	1798	ArgAspArgAspArgGluArgAspArgAspArgGluArgGlySerIleLeuThrSer	1817
Db	17938	ACCACCTCGAC-----	17955
Qy	1818	ThrThrThrValGluHisAlaProIleTyr-----ArgProGlyThrGluGln	1833
Db	17956	ACCTACCTCTTCAACACCAACCACTACTGGCTCGAAAGCACACGCGGGTCCGGATCC	18015
Qy	1834	SerSerGlySer---SerGlySerSerGlyGlyGlySerSer-----	1848
Db	18016	GGTTCCGGTTCCGGTTCGGGGCGGACGACTCGCGGGGAGGAGAGGTGGAGTCG	18075
Qy	1848	-----	1848
Db	18076	CGGTTCTGGGACGGGTGGCGCGCAGGACTCGTGAACGGTTCGCGACGCTCGCGGTG	18135
Qy	1849	-----SerArgProAlaSerHisAla	1857
Db	18136	CCCCCTCCGCGGCTGGACACGGTGTGCGCGACTCTCCGCTCGCACCGCCACCAA	18195
Qy	1858	HisGln-----	1859
Db	18196	CACGACCAAGCCGCATCAACACCTGGACCTTACCAGGAACCTTGAACCCCTCACCCTC	18255
Qy	1860	-----His-SerProIleSerProArgThrGlnAspAlaLeuGlnArgProse	1876
Db	18256	CCCACCAACCAACCAACCAACCTTGGCTCATCGCCATCCCGGAAACCCAGACCCAC	18315
Qy	1876	rValLeuHisAsnThrGlyMetLysGlyIleThrAlaValGluPro---SerLysPr	1895
Db	18316	C-----ACCCGCATCATCAACATCTCTCAC	18342
Qy	1895	oThrValLeuArgSerThrSerThrSerProValArgProAlaAlaThrPheProPr	1915
Db	18343	AACC-----TCCACCAACGCGATCATCCCCCATCCCTCACCTCAACCAACAC	18393
Qy	1915	oAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluPr	1935
Db	18394	C---ACACCAACCCCAACACCTCCACCAACCCGACCAACCAAGCCCAACCAACCA	18450
Qy	1935	oValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspTh	1955
Db	18451	GGACCCATCACCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	18510
Qy	1955	rGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerPr	1975
Db	18511	CCCC-----ACACACCCACCGCACCTCTCTCAACCTCACCTCACCAACCAAC	18558
Qy	1975	oSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAl	1995
Db	18559	CACA-----CCCAACCCCAACCAACCAACCTCTCTCTCTCTCTCTCTCTCTCTCT	18606
Qy	1995	aArgThrProAlaLysAsnLeuAlaProHisAlaSerPro-----	2009
Db	18607	GCCACCAACCCCAACCAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	18666
Qy	2010	-----AspProProAlaProProAlaSerAspPr	2021
Db	18667	CTGCGCGCACCAACCTCTCTCGAACAACCCCAACCCCAACCCCAACCTCATCGACCTC	18725
Qy	2021	o-----HisArgGlyLysThrGlnSerLysProPheSe	2032
Db	18726	CACCAACCCCAACCCCAACCTCTCTCGAACAACCCCAACCCCAACCCCAACCA	18785
Qy	2032	rIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrProGluG1	2052
Db	18786	CCAAACCACTCGCCATCCGACCAACCCCAACCCCAACCCCAACCCCAACCA	18845
Qy	2052	yValGluProValSerProValSerSerProSer-----LeuThrHi	2066
Db	18846	CCTCACCCCAACCAACCAACCCCAACCCCAACCCCAACCCCAACCTCATCCCGG	18905
Qy	2066	sAspLysGlyLeuProLysHisLeuGluLeuAspLysSerHis-----	2081
Db	18906	CGGAACCGGCGC-CCTCGCCACCTCACCCCAACCTCACCCCAACCCCAACCC	18964
Qy	2082	-----LeuGluGlyGluLeuArgProLysG1	2090
Db	18965	ACACCTCTCTCTCACCGCAACCGGCGCCCAACCCCAACCCCAACCTCACCA	19024
Qy	2090	nProGlyProValLysLeuGlyGlyGluAlaHisLeuProHisLeuArgProLeuPr	2110
Db	19025	CCCAACTCCCAACAAA-----AGGCATCCACCTCACCTCACCTCGG	19069
Qy	2110	oGluSerGlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisG1	2130
Db	19070	ACACGACCAACCAACCACTCCA-----ACNACTCTCT-----CAACACCA	19111
Qy	2130	nArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHi	2150
Db	19112	TCCCCCAACACACCCCTCACAC-----CGTCATCCACACCGCGAGCA	19156
Qy	2150	s-----HisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPh	2166
Db	19157	TCCTCGACGACGCCCTCACCAA-----CCTACCCCAACCCCAACCAACG	19207
Qy	2166	eProGlyAlaSerCysProValLeuAspLeuArgProProSerAspLeuTyrLeuPr	2186
Db	19208	TCCTCGCGCCAAAGCCCA-----CAGCGCCACCT-----CC	19240
Qy	2186	oProProAspHisGlyAla-----ProAlaArgGlySerPro-----	2198
Db	19241	TCACCAACTCACCAACACACCCCTCAAGCCTCTCTCTCTCTCTCTCTCTCTCTCTCT	19300
Qy	2199	-----HisSerGluGlyGlyLysArgSerPr	2207
Db	19301	CCACCTTCGGCGCACCCGCGCAAGCAACTACCGCGCAACCAACGCTTACCTCGACGCC	19360
Qy	2207	oGluProAsnLysThrSerValLeuGlyGlyGlyIleGluProValSerPr	2227
Db	19361	TCGCCCA-----	19369
Qy	2227	oProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrAr	2247
Db	19370	ACCGCACACCCACCACTCCCGCCACCAAG-----CATCG	19405
Qy	2247	gAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerG1	2267
Db	19406	CCTGGGGACCTGGCAAG---AAACGACTGGCGACTGGTCAAGTCAGCGCAATCTCC	19462
Qy	2267	nProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLy	2287
Db	19463	GCCGCGCGGGA-TGTTGCGCATGCGCGCGAGTGGCGGTACAGCTGTGTGCGCGCG	19521
Qy	2287	sGlnGluIleAsnLysLeuAsnThrHis-----AsnArgAsnGluProG1	2303
Db	19522	ATCGCGACGCGCGCGAGTCTCTCTCGCGCATATCGACTGGAGAAATTGGGACCG	19581
Qy	2303	uTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThr--	2322
Db	19582	GTTCCTCTCC-AGCAAGTCGCTGCTCTGCTCGAGGACCTTCCCGACGACGAGNACTGA	19640

JOURNAL

Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical Sciences, Kitasato University, Microbial Chemistry; 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan
(E-mail: ikeda@pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242, Fax: +81-3-3444-6197)

FEATURES
source

Location/Qualifiers
1. 64957
/organism="Streptomyces avermitilis"
/mol type="genomic DNA"
/db xref="taxon:33903"
101..12019
/genes="aveA1"
101..12019
/genes="aveA1"
/notes="contains a loading module, Ave module 1 and module 2
multifunctional polyketide synthase"

gene

101..12019

CDS

Qy	2323	-----GlyLeuMetThrTy	2327
Db	19641	GGAGCGCGCAGTACCGTTGAGCAGACGAGACAAACCTCGCGCAACTCTCTCATGGG	19700
Qy	2327	rArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleLea	2347
Db	19701	TCGGTTCAGTTCGCGAGCAAGAGAGCTG-----CTCAGCCTCGTCGCATCCACTC	19754
Qy	2347	gLyAlaLeuMetGlyLysTyraPaspGlnTTPgluLusSerProProLeuSerAlaAenAl	2367
Db	19755	CGCGGAGTCTCGCGCGC---GACGACTCGAGGCCATCCGCCCGG-TCGGCTGTCA	19810
Qy	2367	aPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAAs	2387
Db	19811	GGGATCTAGGTTTCGCTCGCTCGCGCGGTGGAGCTTCGCAACCACTCCGACACAGA	19870
Qy	2387	pGly-----ArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLy	2404
Db	19871	CGGAGCTGGCTCTCGCGACACTCTCGTCTTCGATTACCCAGCCCAACCAAGCTCGCC	19930
Qy	2404	sValSerGlyArgProSerArgLysAlaLysSerProAla-----	2418
Db	19931	AAITCTGCTCT-CCGAGATCGCG---AGTTCAGCCCGACAACTCAACTCGCCTTCG	19986
Qy	2419	----ProGlyLeuAlaSerGlyAspArgProProSerValSer---SerValHisSerG1	2436
Db	19987	CGACCCCGCGCAGACTCGATGACCGCATCGGCATCGTGGCATGGCTGCTGCTCCCC	20046
Qy	2436	uGlyAspCysAsnArgThr-----ProLeuThrAsnArgVa	2449
Db	20047	GGCGGATGACCTCGCGCGACGACTTCTGGGATCTGATCTCTCCGACGAGCGCGCATC	20106
Qy	2449	lTrpGluAspArgProSerSerAlaGly-----SerThrProPheProTy	2464
Db	20107	GGCGGATTCGCCACCGCGCGCTGGGACTGGACAGCTCTACGACCCCGACCCCGAC	20166
Qy	2464	rAsnProLeu-----	2468
Db	20167	CACCCCGGACCTGCTACACCCGAAACCGCGGATTCCTTACGACGCGGCCACTTCGAC	20226
Qy	2468	eMetArgLeuGlnAlaGlyValMetAlaSerProProPro-----ProGlyLeuProAl	2486
Db	20227	GCGGAATTCCTTCGCGATCAGCCCCCGCGAAGCCCTCGCCATGGACCCCGACGACTC	20286
Qy	2486	aGlySerGlyProLeuAlaGlyProHisHisAlaTrpAspGluGluProLysPro	2504
Db	20287	CTCCTCGAAACCGCTCGGAAACCATCGAACCGCGGATCAACCCCGACCCC	20341
RESULT	78		
LOCUS	AB032367	64957 bp	DNA linear BCT 18-SEP-1999
DEFINITION	Streptomyces avermitilis polyketide synthase gene cluster (aveA1, aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.		
ACCESSION	AB032367		
VERSION	AB032367.1	GI:5902890	
KEYWORDS	AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES 4; type I polyketide synthase AVES 3; type I polyketide synthase AVES 2; type I polyketide synthase AVES 1.		
SOURCE	Streptomyces avermitilis		
ORGANISM	Streptomyces avermitilis Bacteria; Actinobacteria; Actinomycetaceae; Streptomyces.		
REFERENCE	1 (bases 1 to 64957)		
AUTHORS	Ikeda, H., Nonomiya, T., Usami, M., Ohta, T. and Omura, S.		
TITLE	Organization of the biosynthetic gene cluster for the polyketide anthelmintic macroide avermectin in Streptomyces avermitilis		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)		
MEDLINE	99380548		
PUBMED	10449723		
REFERENCE	2 (bases 1 to 64957)		
AUTHORS	Ikeda, H., Nonomiya, T., Usami, M., Ohta, T. and Omura, S.		
TITLE	Direct Submission		

translation="MORMDGGEPRAAGVGLVADADGCVVVFVFGCGQPQPMGR
ELLADSDVFRSVCAPAPYVVDWSVEQLRSDPADGLDRVDVVDVQVTLFAVMISL
AALMRQGVPCAVLHGLSIEIAAHSVSGLSADARVVLMSQAQTLTLAGTALYS
VAATPDELLPRIAPMTEDNPALVAVAAPRSTVVGAREAVADLVADLTAAQVTR
MIPVDPLHSLMVAIEERVVGSLPTTPRPSRIPTHSSVTGRLDRELDAAWYRN
MSTVRPEAPARLLLOQPKTFVEMSPHVLWGLQELAPDLGDTGTADTVMGTLR
RGCTUDHLFSLASQURGHGETATTVLARLTALSPQTQOQLDLDLDRHAHTWALND
DGNERTASDGPSAFHILGFDVSMVGNLNRSLKATGLRPLVTLTFDHTTAAVAAR
LRTAALHIDEDTAPVDPSPSGHGTAAADDPITAIIGMACRFGVGRSKDLMELAS
GGAIQGFPTDGTWTEQORHAQDPTQGTFFPQGGFLHDAHFADGAFSGPREALA
MDPQRLLETLSWEAFERAGIDPLSVRSRTGVFAGALSFDYGRMDTASSEGAADVE
GHLLCTTGSLSGLIAYSGLEGPATVDTGCSASLVTLHLACSLRSSECTALAG
GVSMTLGHFISFQRGLSDGRCAVSAADAGTGWGEGVGLLVERLSDAVRLGH
RVLAIVRGSAVNDGASNGLTAPNGPAQERVIRQALANAGLSVADVVDVVGHTGTTL
VDIEAQLALATYQAGRAGPLWGLSKNIGHTMAAGVGGVIMVMALREGLVPT
LHVDPKSPGVLEGAGAVRLTEAVPWPQDAAGRLRAGVSGFGTGGTNAHVILEEAP
AGVCAGGVGLEGAGPLAISVAESVAPVAESVAPVAPVAPVAPVAPVAPVAPVAPV
LRAQALRQYAVRPDVSLADVAGLACGAVLEHRAVLAADRELVGLGALAG
EDRRVTTHRRDAGDADWERADVQVLFVSMVLSAALMRSYGLEPDAVLGHSGEIA
VMSVDDIURDAGDADWERADVQVLFVSMVLSAALMRSYGLEPDAVLGHSGEIA
AHVCGALSKDAKTAVALRSLAALAVRGGRGMASVPLPAQVEQLIGERWAGRLWA
AVNGPRSTAVSGDAEAVDLAYCAGTGVARRIPVDYASHCHPQVOPRELELLELD
ISPQSGVPFSTVEGTWLTTLDAAYWRNLHQPVRFSDAVQALADCHRVFVEVS
PHPLPALEDITTEDAEDVTAIGSLRGNNDTRRELTAHAHTTTCIGCTTTWHHY
THHTPHPHTHLDLFTYFHOHWLESQPGAGSGAGSGAGSGAGSGAGSGAGTAGTGA
EVESRWDVAVARQDLETVAITLAVPPSAGLTVVPALSAWHRHQHOARINTWYQET
WKPLTPTTHQPHQTLIAIPETQTHPHITNLTNLHHGHTPIPLTLNHTHTNQH
LHHTLHTRQQAQNTHTGATGLSLDLDELPHPHHTPTGTLNLTTLTQHTHTPT
PPTPLVATNTTTHNPDLTHPTQATWGLARTLLEHPTHTAGIDILPTTPTPT
LQHLTQTLQPHQTOIARITCTHTTRELTPTLTPTHTQPTPTPHGTLITGGTGL
ALHTLHHTLTQHTLTSRTGPHTPHAQLTLTQOQGIHLITITCTDNPDPQL
QUNLTTPQHTLTVITHTAGILDDATLNTLPTQANNVLRASAKHSLHOLTQHTPL
TAFVLYSAAATFGAPQAAANAYDALAHRTHTLHPATSIAMGTWQGLADS
DKARVLDRRGRFMPSPALATAVQAIADTERPVYVADIADMSKTEHTSDGLVSA
ARERFAVORPPTPASPILKTAHOTSDORAALELVROHVAALVLRHADKATAPDS
FRALGDSLTAVEFNLLIKATGLRPLVSLVFDHTPTAKLAVHLQGLRQTLTAESAPS
AAVATAESVTEPIAIVMACRFPFGVTADDLISSEQDAIGFQPTDRGDLDTL
YDPPDPHTGTCYTRNGQFLYDAGHFAFGISPREALAMPQRLLELTAWETIEHA
GINPHTLHGTCTYNGQDYALRVHAGQSTGDFALTGTAGSVISGRISVTFFH
GPVSVDTTACSSSLVALHACOLRAGECSMALAGVTVMSPPAFVFFORGLAAD
GKCPASAAADGTGCGVGMILLVERLSAHRNGHVLAVVRGSANODCASNGLTAP
NGFSQORVIRQALANAGLSAGDVDAVEAGTGTGLDPTAQAALLATYQDRAEGEPL
WLGVSXNSNGHTQAAAGVAGVIMVLRHGLLPRTLHVDEPSPHGDVMSAGAVQLELLE
TVPWPGCEGLRRRAGVSFGVSGTNAHVILEEAPADDPVGGPPAGDAGDGADEAAG
SPGVWPLVSAKSQPALQAQALHAHLTDHPGLDLADVGYTLARHARVDFDHRATLIA
ADRTFLQALQALAAAGEPHPAVTHSSAPGCTGTGEAGKTAFICSGQGTORPGWAHL
YHTHPVFAALANDICTHLDHPLHPLLLPLLTONDNDNEAALLQOTRYAQPALFAQ
VALHLLTGYHITPHYAGSLGEITAAHLGILTLTDLTTLTQTLTQTLTQTLTQTLT
MTTLHTTPHITLTHLTAHENDLAAINTPTSLVISTGTHPTVQHIITLTCQQQGLTKT
LPTNHAHPSHTNPILNQLHQTTLTHTPHPTLITANTPPDQLTTPHYWTOARNT
VDYATTTQTLHQHGVTTYIELGPDNTLTTLTHNLNPNPTTTTLTTLTHPHHPTQHLIT

D	b	11634	ACCTCGGCTTCAGATCCCTCACCAGCGGTGGAACTTACGCACAACCGGCTCTCCCGCACCAACCG	11693
Q	y	202	GluGluaAlaLysProGluProGluLysProValSerProPro-----ProIle 219 ::: :::	
D	b	11694	GACTCCGCCCTCCCACCACTCGCTTCGACCAACCCCAACCACTCCACCCACC	11753
Q	y	220	GluSerLysHisArgSerLeuValGlnIleIleTyraSpGluAenArg-LysLysAlaGI 239 :::	
D	b	11754	ACCTCCACACACAACTCC-----AGCCAACAACCGGACACAGCTGT-- 11793	
Q	y	239	wAlaAlaHiAargIleLeuGluGlyLeuGlyProGlnValGluLeu---ProLeuTyraS 258 :::	
D	b	11794	-----CGCCCCGTGTGGCGAGTGTCGACAAACTCGAATCCGCCCTCTC----- 11838	
Q	y	258	nGlnProSerAspThrArgGln-----TyrHi 267 :::	
D	b	11839	-CGCCCTCGAAAAACCGACAGCGCCAGCAAAGAGTCAACCTCGCGCTCAAGTCACTCA 11897	
Q	y	267	sGluAenIleLysIleAenGlnAlaMetArgLysLysLeuIleLeuTyraPheLysArgAr 287 ::: :::	
D	b	11898	TGTTGAGTGGAAAG-CACCCACGATCCGACAGCCGAAGAAGCGCTGATGACGAGAGAAG 11956	
Q	y	287	gaEnHisaalargLysGlnTrpLysGlnLysPhe----- 298 ::: :::	
D	b	11957	TTTCACT-----CGGCAACAGAGGTGAGATTTCAAATTAATTGACACGACTCGGC 12010	
Q	y	299	-CySGlnArgTyraSpGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAs 318 	
D	b	12011	CTGTCTGAACCGGACGCTGCCACTCCGC-----CGTATCCGCTG 12052	
Q	y	318	nAenProArgArg-----AlaLysGluSerLysValArgGlnTyraGI 334 :::	
D	b	12053	-GGCCCTCTAGACGTGAATGCAATGGCGAATGAAGCAAGCTCTCGAATACCTCAA 12111	
Q	y	334	uLysGlnPheProGluIleArgLysGlnArg-----GluLeuGlnGluArgMe 350 ::: :::	
D	b	12112	GCGGCTCACTGCGGACGTGACCGCACTCGCGCTCGCTGTACGAGGTGTGTCGAGGTGA 12171	
Q	y	350	tGlnSerArgValGlyGlnArgGly-----SerGlyLeuSerMetSerAl 365 :::	
D	b	12172	GCAGGACCATCGCATGTCGGGTGGGTGGCTGTCTTACCGAGCGGGCGAGCTACCC 12231	
Q	y	365	aAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGlu----- 383 :::	
D	b	12232	CACGCGACTGTGGCATCTCTCAAGTCCACAGCGGACGCTATCGGGGAGTTCCCGACCGA 12291	
Q	y	384	-----AsnLeuGluLysGlnMetArgGlnLeuAlaValIleProMetLeuTy 400 :::	
D	b	12292	CCGTGGATGAACCTGGAGCAG-----CTCTA 12318	
Q	y	400	rApAlaAspGlnGlnArg-----IleLysPheIleAenMetAsnGlyLeuMetAlaAs 418 	
D	b	12319	CGACCGGACCCCGACCGCTCAGAACCATGTTACACGCGAGCGGGAGGTTTCTATGA 12378	
Q	y	418	p-----ProMetLysValTyraSpArgGlnValMetAsnMetTr 432 :::	
D	b	12379	CGCGGGCGACTTCGACCGCGGTCTTCGAGTTGTCAACCGGTGAGCGGTGCGAATGGA 12438	
Q	y	432	pSerGluGlnGlu-----LysGlnThrPheArgGluLysPheMetGln----- 446 :::	
D	b	12439	CCGCGACGACGCTGTGCTCGAAAACCACTTGGGAAACGTTCCGAGCGGGCGGAATCGA 12498	
Q	y	447	-----HisProLysAsnPh 451 :::	
D	b	12499	CCGAGGTCCATCGCGGAACCGGACCGGGGTTTTGTTGGGGATCAATCCGAGGAGCTA 12558	
Q	y	451	e-----GlyLeuIleAlaSerPheLeuGluArg-----Ly 461 	
D	b	12559	CACCAACCGGATAAC-ACATCAGCCCTCAAAACGAGTCGAGGGCTACCTGCTCACTGGCA 12617	
Q	y	461	sThrValAlaGluCysValLeuTyraTyraLeu-----ThrLysLysAsnGI 477 :::	
D	b	12618	GCSCGGCAACGATTCGCTCAGCGGCTATCTCTACCACTTCGGGCTCGAAGGCCCTGCGA 12677	

736 sSerGly-----ProAlaThrValAenAenSerSerAspThrGluSerI 751
 13676 GTCCGGTACACCTCCGCCACGCGCCGGCGGTTCGACACCGCGCCACCTCATCGCC 13735
 751 ePro-----SerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysPr 769
 13736 GCCGACCGGACACCTTCCTGCAAGACATCCAGGACTCCGCGGACGCGGCAACCCACCC 13795
 769 oProAlaThrLeuGlyAlaAsp-----GlyProProProGlyProPro----- 783
 13796 GCCTGTCATCCACAGCAGCGCCCGACGGCGGACCGGACCGGAGCGCCGAGGAGCCAGGAAGACC 13855
 784 -----ThrProProArgThrSerArgAlaProIleGluProThr----- 797
 13856 GCATTATCTGCTCCGACAGGCGACCCCAACGCGCGCGGATGCGCCACGCGCTTACCAC 13915
 798 -----ProAlaSerGluAlaThrGlyAlaPro----- 806
 13916 ACCCACCCCGTTCGCGCGCGACTCAACGACATCTGCACCCACCTCGACCCCGCCACCTC 13975
 807 -ThrProProAlaProProSerProSerAlaProProProValProValProLysGluI 826
 13976 GACCACCCCTCTCCCTCTCTCCAGGACCGCCCAACACCCAGGACACCCACCCCTC 14035
 826 uLysGluGluGluThrAlaAlaAlaProProValGluGluGluGluGluGluGluGlu 846
 14036 GAGAGAGCGCGGCGACTCTCCAGCAGACCGGCTACGCCGCGCGCGCTTCTCGCTTC 14095
 846 oAlaAlaGluGluLeuAlaValAsp-----ThrGlyLysAlaGluProValLysSerG 865
 14096 CAGGTGCGCTCCACCGCTCTCCAGCAGCGGTACACATCACCACCCCGCTACTAGCC 14155
 865 uCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaGluAl 885
 14156 GGACACTCTCTCGGGAATCAGCGCGCCACCTCGCGGCACTCTCACCCTCACCGAC 14215
 885 aThrAlaGluGlyAlaLeuLysAlaGluLysLys----- 896
 14216 GCCACACCTCATCACCAACGCGCCACCTCATGCAACACGATGCCCCCGGACCATG 14275
 897 -GluGlySerGlyArgAlaThrAlaLysSerSerGlyAlaProGlnAspSerAs 916
 14276 ACCACCTCCACACACCCCGCCACATCACCACCTCAGCGCCCGCAACCAAGAC 14335
 916 pSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyLysLysAsnAr 936
 14336 CTGCGCATCGCGCCATCAACACCCCGCTCTCTCGTCAACGCGCA----- 14384
 936 gLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSe 956
 14385 -----CCCCCGACACCGCTCAACACATCACCA-----CCTCTGCCAACAA 14428
 956 rProGlnLysProLeuAspLeuLysGlnLysGlnArgAlaAlaAlaAlaPro-----Pr 975
 14429 GGCATCAAAACCAACCTCCCGACCAACCGCTTCCACTCCCGCCCGCCACCAACCC 14488
 975 oLleGlnValThrLysValHisGluPro-----ProArgGluAspAlaAlaProThrLysPr 994
 14489 ATCCTCAACCACTCCACGACACCAACCCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 14548
 994 oAlaProAlaProProProProGlnAsnLeuGlnProGluSerAspAlaProGlnG 1014
 14549 CTATCACCACCAACCCCGACCGGACCACTCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 14608
 1014 nProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAl 1034
 14609 GCCCGCAACA-----CCGTCGACTACGCGACCAACCCACCAACCCCTCCACCA 14656
 1034 aPeAlaAlaGluAlaGlnLys-----LeuProGlyAspPro-----ProCy 1048
 14657 CACGGCTCACCCTCATCGAAGTCCGACCGGACCGACACACCTCACCACCTCACCAC 14716
 1048 sTrpThrSerGlyLeuProPheProValPro-----ProArgGluValI 1063

14717 GACAACTCCCAACACCCCAACCCCAACCTCACCCTCACCCTCACCCTCACCCTCACCCTC 14776
 1063 eLysAlaSer-ProHisAlaProAspProSerAlaPheSerTyAlaProProGly---H 1082
 14777 CAACCCACCTCTCTCACCACCTCGCCAAACACCCACCCACCTCGGACCCCGCCACTAC 14836
 1082 isProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProAspProProThrI 1102
 14837 ACCCACCAACA-----CAACCAACCCCAACCCCAACCCCAACCTCGACCTCCCGACC 14887
 1102 leSerAsnPro---ProProLeuIleSer-----SerAlaLysHisProSerV 1117
 14888 TACCCCTTCCAAACACCACTACTGCTCCAAACCCCGCAAGCGGACCGCC---G 14944
 1117 alLeuGluArgGln----- 1122
 14945 TCACGAGCGAAGCGCGTGAGCAAGCCACGACCCCATCAACCCCGCTGCGTGATGCTCTC 15004
 1122 leGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyTySerGlu-HisAla 1141
 15005 GTGGCAAGTCTCCGAGGAGGAGAGAGAGCTGTGGCGCTGTGGCGACCCATGCG 15064
 1142 LysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeu 1161
 15065 GCGCTGTGCTGGCGCATGCGCTCCGAGTATGCTTCGGAACAAGCGCTTCAAAGAG 15124
 1162 AlaProPheSerGlyValLysGlnGlnLeuSerProArgGlyGlnAlaGlyPro--- 1180
 15125 CTGGTGTGATCTCTCCGCGCAATTCAGCTTCGTAATCAGCTGCTGCTGACGTTGAC 15184
 1181 -----ProGluSerLeu-----GlyValProThrAlaGlnGlu---AlaSerVal 1194
 15185 CTGCGCTTCCGCGCCAGCTGATCTTCGATTACCCCACTCCGATGGCGCTTTCGACGTC 15244
 1195 LeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyLysProSer 1214
 15245 CTGCGGCGCGCATGCTGCGAGCGGACACAGGC-----ACGACCATCTGCTGCGGCTA 15298
 1215 ThrArgValProSerAspSerAlaIleThr-----TyrArgGly 1227
 15299 ACTGCGTCTCCGCGACGACGCGGATCGCATCGTGGCATGCGCTGCTGCTGCTGCTGCT 15358
 1228 SerIleThr----- 1230
 15359 GATGTACGACGCGTGCATGATCTCTGCGAGTGTGAGTGTGCGCATGACCGGATCGCG 15418
 1230 ----- 1230
 15419 GGATTCGCGACGAAACCGTGGTGGGACCTCGACACGCTGTACAAACCGGACCGGACCC 15478
 1231 HisGlyThr----- 1233
 15479 CACGGAACGACTACACCGGAGCGGATTCCTTTACGACGACGAGCAATTCGATGCC 15538
 1234 -----ProAlaAspValLeu----- 1238
 15539 GACTTCTTGGTATCATCTGCGTGGAGCACTGGCATGGACCGGACGCGGCTGCTG 15598
 1239 -----TyrLysGly 1241
 15599 CTGGAACACGCTGGGAGACATCGAAGCGCTGCATCAACCCCGACGACCTCCGTTGGC 15658
 1242 ThrIleThrArgIleIleGly-----GluAspSerProSerArg----- 1254
 15659 ACACCAACCGGCGTCTTCCGCGGCTGACCTACCGAGCTACCGCGCGGCTTCCACCA 15718
 1255 -----LeuAspArgGlyArgGluAspSerLeuProLysGly 1266
 15719 GCTCGGACGAGTTCGAGGGTATCTCGGACGGAAGCGAGCAGTATCGCTCGCTCGGT 15778
 1267 HisValIleThrGlu---GlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSer 1285

Db 15779 CGTCTCGCTACGCTCTCGCGCTCGAAGGTCGGCCCTCACAGTCGAC----- 15826
Qy 1286 ValThrGlnCysSerLysGluAspGlyArgSerSerGlyProProHisGluThrAla 1305
Db 15827 ---ACTGCTGCTCTCTGCTCCT--GTCGCTCTGCA-----CTGGCCTGTACGGCGCT 15876
Qy 1306 AlaProLysArgThrTyr-----AspMetMetGluGly 1316
Db 15877 GCGTCTCGGCGAGTGTTCATGCGCTCGCGGTGGGTGTACGGGTGATGTCAACCCCGGC 15936
Qy 1317 ArgValGlyArgAlaIleSerSerAla-----SerIleGluGlyLeu 1330
Db 15937 CGGGTTCGTGAGTTTCGCGGACAGCGGGGCTCGCGGTGACGCGGGTGCAGAGCGGT 15996
Qy 1331 MetGly----- 1332
Db 15997 CTCGGACGGCTGACGGCACCGGCTGGGGTGGGGTGTGCGAATGCTGTGTTGGAGCG 16056
Qy 1332 ----- 1332
Db 16057 GCTGTGCGACGCGCGGCTCGGTCTCGGTACCGAATCTCGCGTGTGTGCGAGTGGCGT 16116
Qy 1333 -----ArgAlaIlePro----- 1336
Db 16117 CAATCAGGACGCTCGAGCAACAGGCTGACGCGGCCCAACGCGGCTGCCAGGAGCGTGT 16176
Qy 1337 -----ProGluArgHis-----Ser 1341
Db 16177 CATCCGCTGGCCCTGCGCAACGCGGACCTGACCCCGCGCAGCTCGATCGGTGGAGGC 16236
Qy 1342 ProHisLysLeuLysGluGlnHisIle---ArgGlySerIleThrGlnGlyPro 1360
Db 16237 CCACGGCAC-----CGGCACCACTTTGGCGGACCGATCGAGGCCCGGCTCCT 16287
Qy 1361 ArgSerTyrValGluAlaGlnGluAspTyrLeuArgGluAlaLysLeuLeuLysArg 1380
Db 16288 CGCCACCTA-----CGG 16299
Qy 1381 GluGlyThrProProProProSerArgAspLeuThrGluAlaTyrLysThrGln 1400
Db 16300 ACAGGACGCGCGGCAACGACCGCTGTGGCTCGATGAGTTCGACATCGGCCA 16359
Qy 1401 AlaLeuGlyProLeuLysLeuLysProAlaHisGlu----- 1412
Db 16360 CGCGCAGGCTCGCGAGGTGTGGCGGGTTCATCAAGATGTGTGCTGCGGAATGG 16419
Qy 1413 -----GlyLeuVal----- 1415
Db 16420 TCTGCTCGCGGACGTTGATGTGTGATGACCGCTCGCGCATGTGGACTGTCCGCGG 16479
Qy 1416 -----AlaThrValLysGluAla 1421
Db 16480 GCGGTGTCAGTGTGACGAGAGCGTGCCTGCGCGCGGGAGGGGGTGGCGG 16539
Qy 1422 GlyArgSerIle-----HisGluIleProArgGluGluLeuArgHis 1435
Db 16540 GGCAGGAGTGTATCTGTCGCGGTGAGCGGCACCAACGCCCGCTCATCTCGAAGAAGC 16599
Qy 1436 Thr---ProGluLeuProLeu----- 1441
Db 16600 ACCGCCCAACAATCCGTCGACACACCGCGCGAGCGCCCGGGAGAGCAGCGCG 16659
Qy 1442 -----AlaProArgProLeuLysGluGlySerIleThr 1452
Db 16660 CGAGGATGTTCCGGGGAAGGCGCGGAGCGCGGTACCGCGGGGAGC----- 16713
Qy 1453 GlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAsp 1472
Db 16714 GACTGTCCTGCTCGCGCAGTCCAGGGGTGTGCGGTGTGCTGTCGCGCAAGTCGCA 16773
Qy 1473 ValArgSerLeuIleGlySerProGlyArgThrPhePro----- 1486
Db 16774 GCGGCGCTGCGGCGCCAGCGCCGAGGCTGTGACGCGCCCACTTCACCGACCAACCCCGGCT 16833

Qy 1487 -----ValHisProLeuAspValMetAlaAspAlaArgAlaLeuGlu 1500
Db 16834 CGACCTCGCCGACGCTCGGGTACACCT-----CGCCACGCGCG 16872
Qy 1501 ArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSerSer----- 1518
Db 16873 CGCGG-TGTTTCGACCAACCGCGCCACCTTCATCGCGCGCGCGGACACCTTCCTGCAAG 16931
Qy 1519 -----GlyGlySerIleAlaArgGlyAlaProVal 1528
Db 16932 CACTCCAGGACCTCGCGCGAGGGAACCCACCCCGCGCTCATCCACAGCAGCGCCCGAG 16991
Qy 1529 IleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAla 1548
Db 16992 GCGGACCGGAGCGGGAGCGCGAGAAAGACCGATTCTATCTGCTCCGAGCAGGGCA 17051
Qy 1549 ProPheAlaGlyHisLeuProArg----- 1556
Db 17052 CCAACGCGCGGATGGCCCGCTTACCACACCCACCGCTCTTGGCCCGCGCAC 17111
Qy 1557 -----GlySerProValThr 1561
Db 17112 TCACAGCATCTGACCCACTCGACCCCACTCGACCCCTCTCTCCCTCTCTCA 17171
Qy 1562 MetArgGluProThrProArgLeu----- 1569
Db 17172 CCAGGACCCCAACACCCAGGACACCAACCCCTCGAAGAGCGCGCTCTCCAGC 17231
Qy 1570 -----GlnGluGlySerLeuSerSerSerLysAlaSerGln 1581
Db 17232 AGACCCCGTAGCGCGCGCTCTTTCGCTTCCAGGTTCGCGCTCCACCGCTCTCA 17291
Qy 1582 AspArgLysLeuThrSerThrProArgGluIle-----AlaLysSerPro 1596
Db 17292 CGACGGCTACCATCACCCCACTACTAGCGGAGACACTCTCTCGGGAATACCG 17351
Qy 1597 -----HisSerThrValPro-GluHisProHisProIleSerPr 1610
Db 17352 CGGCCACCTCGCGGATCTCTCACCTCACCGGACCCACCTCATCACCAACGCG 17411
Qy 1610 oTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLe 1630
Db 17412 CCACCTCTCGAACAACCTCGCGCGGACCACTTACACCCCTCCACACCCCGCCACC 17471
Qy 1630 uAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaIleTyr-- 1649
Db 17472 ACATCACCACCACTCACCGCCACGAAACGACCTCGCGCATCGCGCCATCAACACC 17531
Qy 1650 -TyrLeuProArgHisLeuAla----- 1656
Db 17532 CCACCTCTCTGTCATCAGCGGACCCCGCCACACCGTCCACACATCACCCCTCTGCC 17591
Qy 1656 ----- 1656
Db 17592 AACACAAGGATCAAAACCAAAACCTCCCGCCACCAAAACGCTTCACCTCCCGCCACA 17651
Qy 1657 -----ProAsnProThr-----TyrProHisLeu-TyrProProT 1668
Db 17652 CCAACCCCATCTCAACCAACTCCACGAGCACCCCAACCCCTACCTACCCACCCCC 17711
Qy 1668 yr-----LeuIleArgGlyTyrProAspThrAlaAlaLeu----- 1679
Db 17712 ACACCCCTCATCAGCCCAACACCCCGCCAGCAACTCTCTCACCCCTACTTGA 17771
Qy 1680 -----GluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMethHisHis- 1697
Db 17772 CCAACAAGCGCGCAACACCGTC-----GACTACGCGCACCAACCCCAACCTCCACC 17825
Qy 1698 AsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGlu 1717
Db 17826 AACAGCGGCTACCACTATCATCGAATCGGAC-----CGGACAACA 17867

Db 19682 GTTCTCTCC-AGCAAGTCGTGGTCTTGTCTGAGGACCTTCCCGAGCAGCAGGAACTGA 19740
Qy 2323 -----GlyLeuMetThry 2327
Db 19741 GGAGCGCGGAGTACCTGTGAGCAGACGAGAGCAGACAACTCCGGCACTCTCATCGG 19800
Qy 2327 rArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaLeuLea 2347
Db 19801 TCGGTCACTGTCCGAGCAGGAAGAAGACTG-----CTCAGCTCTCGCGCATCCACTC 19854
Qy 2347 gLyAlaLeuMetGlyLysTyAspGlnTrpGluLysSerProLeuSerAlaAsnAl 2367
Db 19855 CGCGGAGTGTCTGGCGC-----GAGACTCGAGCCATCCCGCCCG-TCGGGTGTCA 19910
Qy 2367 aPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAs 2387
Db 19911 GGGATCTAGGTTCCGACTCGTTCGGCGGTGGAGCTTCGACCACTCGCAGCAGAGA 19970
Qy 2387 pGly-----ArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLy 2404
Db 19971 CGGAGCTGGCTCTCGCAGCACTCTGCTCTCGATTACCCAGCCAGCAGCTCGCC 20030
Qy 2404 sValSerGlyArgProSerSerArgLysAlaLysSerProAla----- 2418
Db 20031 AATTCTGCTCT-CCGAGATCGCG-----AGTTCAGCCCGCAGCACTCAACTCCGCTTCG 20086
Qy 2419 ----ProGlyLeuAlaSerGlyAspArgProProSerValSer---SerValHisSerG 2436
Db 20087 CGACCCCGGAGAGCTCGATGAGCCGATGCCATCTGTGGCATGGCTGTGCTTCCC 20146
Qy 2436 uGlyAspCysAsnArgThr-----ProLeuThrAsnArgVa 2449
Db 20147 GCGGAGTGCACCTCGCGGAGCACTTCTGGATCTGATCTCTCCGAGCAGGAGCGCATC 20206
Qy 2449 lTrpGluAspArgProSerSerAlaGly-----SerThrProPheProTy 2464
Db 20207 GCGGATTCCTCCACCGGCGGTGGACCTGGACAGCTCTACGACCCGACCCCGAC 20266
Qy 2464 rAsnProLeu----- 2468
Db 20267 CACCCGCGCCTGTCTACACCCGAGCGGATCTCTTACGAGCAGGCGCCTTCGAC 20326
Qy 2468 eMetArgLeuGlnAlaGlyValMetAlaSerProProPro-----ProGlyLeuProAl 2486
Db 20327 GCCGAATCTTCGCGCATCAGCCCGGAGAGCCCTCGCATGAGCCCGCAGCAAGCACTC 20386
Qy 2486 aGlySerGlyProLeuAlaGlyProHisHisAlaTrpAspGluProLysPro 2504
Db 20387 CTCTCGAACCCTCGGAAACCATCGAAGCGCGCATCAACCCCGACCC 20441

RESULT 79
LOCUS AP005024 302675 bp DNA linear BCT 10-MAY-2003
DEFINITION Streptomyces avermitilis genomic DNA, complete genome, section 4/30.
ACCESSION AP005024 BA000030
VERSION AP005024.1 GI:29604389
KEYWORDS Streptomyces avermitilis MA-4680
SOURCE Streptomyces avermitilis MA-4680
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1
AUTHORS Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C., Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T., Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
TITLE Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
MEDLINE 21477403
PUBMED 11572948
REFERENCE 2

AUTHORS Ikeda,H., Ishikawa,J., Hanamoto,A., Shinose,M., Kikuchi,H., Shiba,T., Sakaki,Y., Hattori,M. and Omura,S.
TITLE Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis
JOURNAL Nat. Biotechnol. 21 (5), 526-531 (2003)
MEDLINE 22608306
PUBMED 12692562
REFERENCE 3 (bases 1 to 302675)
AUTHORS Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C., Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T., Kishida,N., Director-General of Biotechnology Center, Shiba,T., Sakaki,Y. and Hattori,M.
TITLE Direct Submission
JOURNAL Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center, 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
COMMENT (E-mail:bio@nitech.go.jp, URL:http://www.bio.nitech.go.jp/, Tel:81-3-3481-1933, Fax:81-3-3481-8424)
This work was done in collaboration with Haruo Ikeda(*1), Jun Ishikawa(*2), Akiharu Hanamoto(*3), Chigusa Takahashi(*3), Mayumi Shinose(*3), Hiroshi Horikawa(*4), Hidekazu Nakazawa(*4), Tomomi Osonoe(*4), Norihiro Kishida(*4), Hisashi Kikuchi(*4), Tadayoshi Shiba(*5), Yoshiyuki Sakaki(*6,*7), Masahira Hattori(*1,*7) and Satoshi Omura(*1,*3).
Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa.
*1 Kitasato Institute for Life Sciences, Kitasato University
*2 National Institute of Infectious Diseases
*3 The Kitasato Institute
*4 National Institute of Technology and Evaluation
*5 School of Science, Kitasato University
*6 Institute of Medical Science, University of Tokyo
*7 RIKEN, Genomic Sciences Center
Following url is also available.
http://avermitilis.ls.kitasato-u.ac.jp.
FEATURES
source 1. 302675
/organism="Streptomyces avermitilis MA-4680"
/mol_type="genomic DNA"
/strain="MA-4680"
/db_xref="taxon:227882"
/notes="This strain is also named as strain: ATCC 31267, NCIMB 12804 or NRRL 8165.-synonym: Streptomyces avermectinius"
CDS 12..425
/note="SAV750"
/codon_start=1
/transl_table=11
/product="putative TetR-family transcriptional regulator"
/protein_id="BAC68460.1"
/db_xref="GI:29604390"
/translation="MAELCLDREARTAAQAALAGLAGCTVARNLDAVALAFDSCG PAMSGAMTRPAALRIREALVDGAGFTAIQEAITGLYKAERGLGRVAETVDRTPVA LAIVGTAHLMLTWSFGAPDRPHMARLVAALVDS"
1504..2592
/note="SAV751"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68461.1"
/db_xref="GI:29604391"
/translation="MRTVGVEELLVDPETGPKALSTAVLARAEOVDPQDVFEKE LHQMLEFATHPTQDMAALCAEIVRCKEARHAGGCAVALATSPVPSIANN RRYQWAEQYGIAMQQLTCGCHVHVAVSDEGVAVDRIEPLVPLVALSANSPPW QGRDSYSTRVGRWFSAGTFLGSPERVHRQVADRVGVLDDGMYVDFDLK SHRYPTVEIRVADVCLHPDTAQIATLARGLVESAARDWRAGREPDSVLLRLAAW QARAGLSGELLHPVTMRIRAESVVRALLHGVGDALADGLDLAHEAVAEALLERGN GARVQELLERTGLRDTVTECVROTG"
complement(3011..3400)
/note="SAV752"
/codon_start=1
/transl_table=11
/product="hypothetical protein"

```

/protein_id="BAC68462.1"
/db_xref="GI:29604392"
/translation="MOITASTVSLTVADVAASQAFITAHGLGVTVQAAADGFFASLSTRAD
ADIVLLARGQVLPQRDRHAAGLILATFADGLHDQEKRLREAGVEITWPLREEP
WGERLFOITPNGVVIQFVEWATSDDT"
complement(3457..3648)
/notes="SAV753"
/codon_start=1
/codon_table=1
/product="hypothetical protein"
/protein_id="BAC68463.1"
/db_xref="GI:29604393"
/translation="MFELDRITTVSALTIGLGEVPAGDVKDPAPRGYDLRLRLAADRA
VRGRPRVVRNIPRRGY"
3755..4408
/notes="SAV754"
/codon_start=1
/codon_table=1
/product="hypothetical protein"
/protein_id="BAC68464.1"
/db_xref="GI:29604394"
/translation="MGTEITSPAQVGLRESKKHETQLISDHATRLFLAQGFQOTTI
AEIAARMAKTVTVFARKEDLADHQEFTASLANAVADRPPGESALAAALRLRAF
DSLATQNPVAGSGPAPFARMVADSPTLTACLGILDHQREHRLAQALADATDTTPDDIT
VHIAGLGLTGHVRVDFQRIQDLITLGHNDRIAHTITAEATQAFDLLEPALAHAYATA"
4587..5627
/notes="SAV755"
/codon_start=1
/codon_table=1
/product="putative dehydrogenase"
/protein_id="BAC68465.1"
/db_xref="GI:29604395"
/translation="MISPKRTRTTPVACPATGSPSQATSPSYPAVPTACRSTWT
CRSDIRPAPAPVPRSPQADFADFPALVSRDLSLITPEERHVLRSVLLSDFDP
LVTRAGTADPALRLITERFVRKENTPCGAAPPAPGHPDRMAVSGRPSGDP
APTSSRPASPAGRNPHRGSAVGTWGAFTPEKWLAVLAVLDVAHHITGRAALHRL
AARGNVSAGSLGIDGYMAAYNAKGAVNLTAMALDHAPEVRVNAHGPGFTLHSG
TAGATPDTPVRSFAERIPLGRVAEPDEIASVVAFLAGPDALYDRRSNPRGRREPHRC
HAAHRDVNAS"
complement(6101..8071)
/notes="SAV756"
/codon_start=1
/codon_table=1
/product="hypothetical protein"
/protein_id="BAC68466.1"
/db_xref="GI:29604396"
/translation="NERSKDRPVSLODRARYWFDRTLARBSTGLMGLVITICLAVVV
PVSALLVMTDPSRSLRSGLAAVWRSAETRLRGTVGTPLMLLSALGLVALVC
STLVGVTGLAEALISRGSTVLEQGHAVILGWSQDSTVVGELVAASQSVYRPRRA
VVVLAERDTEREALAAHVPAGRTIRGVSFGSPADSGVLAVSPVASTVLVLPSSG
EPTADAEVRLVLLALGEGTGPVLLAALDDRYRAPRLAAGRTGFTVLEDTVTV
ARLIAQVCRGSLVLRLDITFDAGDEPFLAABATFHGCGFCAALLGHATSGVGLLT
AGERTLIPNPATLVAPGSRVLVLTDDGSAVPEDCHLVPESAIAWAQPPEDAAHL
RLGLGNRAPLVNQLRTRSGVLVDVDRVAPGPTREPESGARVLRVFSQADLS
LETLLGLDLPYDGLVILGPPFGGDPDRDRTLVLALLRLLEERTGEMRVITEL
TDDRNPPLAVPNPGSDVIRGELTGLLQAQONKHLPAVDFELFSADGSTICLRQAG
HYVTRGCTETATVAAADRGECAIGYRRHRRRAIPDPDHGIRLNPDKSERREWSADD
QVVVVAETPEPAAGQDTEHQPGPQPTTRA"
9086..10072
/notes="SAV757"
/codon_start=1
/codon_table=1
/product="hypothetical protein"
/protein_id="BAC68467.1"
/db_xref="GI:29604397"
/translation="MGRRKGGIGIALVTVSMLVGASACGGSGDKANGDARAKSGKFP
ASASPKPKAGPMLLETITPTQTTGVGVAMPISVVFTNPAAKARAARAEVKHMKVS
QPVSGAWEHFGKXADRPKTPWPGTKVKITDAMKGVNNGRYGVHSVTHTFKIC
SDVRADSVTGHMKVRKNGKVVRILSINAGSAQPTWNGTMAVTDKQEKVHMTSCS
GISCEKGSNPYDITLPWDVHLTQSGTIVYITGDPDTFGSGSARGSHGCVHLMSMDA
WFYQGVKQGDPTVITGSPRAKAPADNGYADFNLQMDQLAGSAGEKTTATL"
complement(10889..11494)
/notes="SAV758"

```



```
QY 736 sSerGly-----ProAlaThrValAsnAsnSerSerAspThrGluSeril 751
Db 251392 GTCCGGTACACCTCGCCACGCGCGCGTGTTCGACACCGCGCCACCTTCATCGCC 251451
QY 751 ePro-----SerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysPr 769
Db 251452 GCCGACCGGACACTTCTCGAAGCMTCCAGGCATCGCCGACGGGACCCACCC 251511
QY 769 oProAlaThrLeuGlyAlaAsp-----GlyProProProGlyProPro----- 783
Db 251512 GCCGTATCATCAGCAGCGCCCGAGCGGAGCCCGGAGCGCGGAGGAGCCGAGGAAAGAC 251571
QY 784 -----ThrProProArgArgThrSerArgAlaProIleGluProThr----- 797
Db 251572 GCATTTCATCTGCTCCGACAGGCGACCCAAACGCGCGCGATGCCACCGCCTCTACCAC 251631
QY 798 -----ProAlaSerGluAlaThrGlyAlaPro----- 806
Db 251632 ACCACCCCGTCTTCGCGCGGCACTCAACGACATCTGCACCCACCTCGACCCCGCCCTC 251691
QY 807 -ThrProProProAlaProProSerProSerAlaProProProValProLysGluL 826
Db 251692 GACCACCCCTCTCTCCCTCTCTCACCAGAGCCCAACACCCAGGACACCCACCCCTC 251751
QY 826 uLysGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGluLysProPr 846
Db 251752 GAAGAAGCGCGCGCACTGCTCCAGCAGACCGCTAGCCCGCGCGCTCTTCGCGCTTC 251811
QY 846 oAlaAlaGluGluLeuAlaValAsp---ThrGlyLysAlaGluGluProValLysSerGl 865
Db 251812 CAGTCCGCTCCACCGCTCTCTCCAGCGGCTACACATCACCCCGCCACTACTAGCC 251871
QY 865 uCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAlaGluAlaGluAl 885
Db 251872 GGACACTCTCTCGCGGAAATCAGCGCGCCACCTCGCGCGCATCTCTCACCCCTCACCGAC 251931
QY 885 aThrAlaGluGlyAlaLeuLysAlaGluLysLys----- 896
Db 251932 GCCACCACTTCATCCCAACGCGCCACCTCTATGCAACCATGCCCCCGGACCATG 251991
QY 897 -GluGlyLysSerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAs 916
Db 251992 ACCACCTTCACACCCCGCCACCATCACCATCACCCCTCACCGCCCAAGAAACGAC 252051
QY 916 pSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyLysAspLysAsnAr 936
Db 252052 CTCGCCATCGCGCGCATCAACACCCCGCCACCTCTCTCTCTCAGCGGCA----- 252100
QY 936 gLeuLeuSerProArgProSerLeuLeuThrThrGlyAspProArgAlaAsnAlaSe 956
Db 252101 -----CCCCCGCCACCGTCCAAACATACCA-----CCCTCTGCAACAA 252144
QY 956 rProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaAlaPro---Pr 975
Db 252145 GGCATCAAAACCAAAACCTCCCGCCACCAACACCGCTTCACCTCCCGCCACCAACCC 252204
QY 975 oIleGlnValThrLysValHisGluPro---ProArgGluAspAlaAlaProThrLysPr 994
Db 252205 ATCTCTCAACCACTCTCCAGCAGCACACCCAAACCTCTACCTACCCACCCCGCCACCC 252264
QY 994 oAlaProProAlaProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGl 1014
Db 252265 CTGATCAGCGCAACACCCCGCCAGCCCACTCTCTACCCCGCCCGCTACTGACCCCAACAA 252324
QY 1014 nProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAl 1034
Db 252325 GCGCGCAACA-----CGCTCGACTAGCGCACCAACCAACCCCTCCACCAA 252372
QY 1034 aPheAlaAlaGluAlaGlnLys-----LeuProGlyAspPro---ProCys 1048
Db 252373 CACGCGGTCAACCATCTACATCGAACTCGGACCCCGCAACACCCCTCACCCCTCACCCAC 252432
QY 1048 sTrpThrSerGlyLeuProPheProValPro-----ProArgGluValil 1063
Db 252433 GACAACCTCCCAACACACCCCGCCACCTCACCTCACCTCACCTCACCTCACCTCACCT 252492
QY 1063 eLysAlaSer-ProHisAlaProAspProSerAlaPheSerTyAlaProProGly---H 1082
Db 252493 CAAACCCCACTCTCTCAACAACTCGCCAAACACCCACCACTCGGCACCCCGCCACCTAC 252552
QY 1082 iProLeuProLeuGlyLeuHisAspThrAlaAlaProValLeuProArgProProThrI 1102
Db 252553 ACCCAACCA-----CAACCAACCCCAACCAACCAACCACTCGACCTCCCGCC 252603
QY 1102 lSerAsnPro---ProProLeuLysLys-----SerAlaLysHisProSerV 1117
Db 252604 TACCCCTTCCACACCACTACTGCTTCCAAACCAACCGGCAAGCGGACCGGACCC---G 252660
QY 1117 aLeuGluArgGln----- 1122
Db 252661 TCACCGAGCAAGCGCGTGAGCAAGCCACCGCCCATCAACCCCGCTCGGTGATGTCCTC 252720
QY 1122 leGlyAlaLysSerGlnGlyMetSerValGlnLeuHisValProTyTrSerGlu-HisAla 1141
Db 252721 GTGGGCAAGTCTCCGAGGAGCGAGAGAGAGCTGTTCGCGCTGTGGCGGACCATGCG 252780
QY 1142 LysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeu 1161
Db 252781 GCCGCTGTCTGGCCATGCCACTCCGAAAGTATGTTCCGAAACAAAGCCCTTCAAAGAG 252840
QY 1162 AlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyPro--- 1180
Db 252841 CTGGGTTTGTATCTCTCGCCCAATTGCTTCTGTAATCGACTGCTTCTGACGTGAC 252900
QY 1181 -----ProGluSerLeu-----GlyValProThrAlaGlnGlu---AlaSerVal 1194
Db 252901 CTGCCCTTCCGCGCAGCTGATCTTCGATTACCCACTCCGATGGCGTGTTCGACGTTC 252960
QY 1195 LeuArgGlyThrAlaLeuGlySerValProGlyLysLysLysLysLysLysLysLysLys 1214
Db 252961 CTCCGGCGCGCATCTCGGAGCGGACACAGGC-----ACGACCACTCTGTCGCGCTA 253014
QY 1215 ThrArgValProSerAspSerAlaLysLys-----TyrArgGly 1227
Db 253015 ACTGCGGTCTCCCGCCAGCAGCGCATCGCCATCGTCGCGCATGCGCTGTCGTTACCCCGT 253074
QY 1228 SerIleThr----- 1230
Db 253075 GATGTACGAGCGTCTGATCTCTGGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 253134
QY 1230 ----- 1230
Db 253135 GGATTCGCGAAGCGTGGTGGGACCTCGACACGCTGTACAAACCGGACCGCGGAC 253194
QY 1231 HisGlyThr----- 1233
Db 253195 CACGGAACCACTGATACACCGGAGCGGGGATTCCTTTACGACGCGGCAATTTGATGCC 253254
QY 1234 -----ProAlaAspValLeu----- 1238
Db 253255 GACTTCTCGTATCAGTCCGCTGAGGACCTGGCGATGAGCCCGCAGCAGCGGTGCTG 253314
QY 1239 -----TyrLysGly 1241
Db 253315 CTGGAACAGCGTGGGAGGATCGAACAACGCTGATCAACCCCGACAGCTCCCGTGGC 253374
QY 1242 ThrIleThrArgIleLysLys-----GluAspSerProSerArg----- 1254
Db 253375 ACACCAACCGCGCTCTTCGCGCGGTGACCTACCACTACGACCGCGCGGTTCCTCCACA 253434
QY 1255 -----LeuAspArgGlyArgGluAspSerLeuProLysGly 1266
Db 253435 GCTCCCGCAGGGTTCGAGGGGTATCTCGGCGACGGAAGCGCAGGAGTATCGCTCGGT 253494
QY 1267 HisValIleTyrGlu---GlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSer 1285
```

	Db	253495	CGTGTGCGCTACGCTCTCGGCCTGGAAAGGTTCGGCCCTCACAGTCCGAC-----	253542
	Qy	1286	ValThrGlnCysSerLysGluAspGlyArgSerSerGlyProHisGluThrAla 1305	
	Db	253543	---ACTGCTGCTCTTCGTCCTC--GGTGCTCTGCA-----CCTGGCCTGTCAAGCGCT 253592	
	Qy	1306	AlaProLysArgThrTyR-----AspMetMetGluGly 1316	
	Db	253593	GCGGTCCGGCGAGTGTTCCATGGGCCCTCGCGGTGGCGTCAACGGTGATCAACCCCGGC 253652	
	Qy	1317	ArgValGlyArgAlaIleSerSerAla-----SerIleGluGlyLeu 1330	
	Db	253653	CGGTTCTGTGAGTTTTTCGCGCAGCGGGCCTGCGTGGAGCGGCGGTGAAGCGGT 253712	
	Qy	1331	MetGly-----	1332
	Db	253713	CTCGCAGCGGCTGACGCACCGGCTGGGTGAGGTTGCGAATGCTCTCGTGGTAGCG 253772	
	Qy	1332	-----	1332
	Db	253773	GCTGTGCGACGCGCGGCTCGGTCAACGAATCCTGCGGTGGTGGTGGAGTGCCTCTCGTGGTAGCG 253832	
	Qy	1333	-----ArgAlaIlePro-----	1336
	Db	253833	CAATCAGAACGTCGAGCAACGGGCTGACGGCGCCCAACGGCGCGTCCCAGGAGCGTGT 253892	
	Qy	1337	-----ProGluArgHis-----Ser 1341	
	Db	253893	CATCGGCTGGCCCTGGCCAACGGGACCTGACCCTCCCGCGCATGTCGTCGGTGGTAGCG 253952	
	Qy	1342	ProHisLeuLysGluGlnHisIle---ArgGlySerIleThrGlnGlyIlePro 1360	
	Db	253953	CCACGGCAC-----CGGCACCACCTTGGCGCACCGCATCGAGGCCACGGCCCTCT 254003	
	Qy	1361	ArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLysArg 1380	
	Db	254004	CGCCACCCTA-----CGG 254015	
	Qy	1381	GluGlyThrProProProProProSerArgAspLeuThrGluAlaTyrLysThrGln 1400	
	Db	254016	ACAGCGCCCGCGCAACACCGCTGTGGTGGGTCTCATGAAGTGAACATCGGCCA 254075	
	Qy	1401	AlaLeuGlyProLeuLysLeuLysProAlaHisGlu-----	1412
	Db	254076	CGCGAGGCTCCGACAGTGTGGCGGGGTATCAAGATGGTATGGCGCTCGGGAATCG 254135	
	Qy	1413	-----GlyLeuVal-----	1415
	Db	254136	TCTGCTCGCGGAGCTTGTCATGTGATGAGCCGTGCGCGCATGTGTGGATCGCGGG 254195	
	Qy	1416	-----AlaThrValLysGluAla 1421	
	Db	254196	GCGGTGACGTGTGACGGAGACGGTCCCTGCGCCCGGGGAGGGCGGCTCGCGCG 254255	
	Qy	1422	GlyArgSerIle-----HisGluIleProArgGluGluLeuArgHis 1435	
	Db	254256	GGCAGGAGTGTTCATGTTCCGCGTCAGCGGCACCAACGCCACGTCATCTCTCGAAGACG 254315	
	Qy	1436	Thr---ProGluLeuProLeu-----	1441
	Db	254316	ACCGCGCACAAACATCCGCTCAGACACACCGCCGACGACGCCCGGGAGAGCAGCGCG 254375	
	Qy	1442	-----AlaProArgProLeuLysGluGlySerIleThr 1452	
	Db	254376	CGACGATGTTCCGGGGGAACGGCGCGCAGCAGCGCGGTACCGCGGGGAGC----- 254429	
	Qy	1453	GlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAsp 1472	
	Db	254430	GACTGGTCTGCTCGCGGAGTCCAGGGGTGTGGCCGTGGTGTCTGGCCCAAGTCGCA 254489	
	Qy	1473	ValArgSerLeuIleGlySerProGlyArgThrPheProPro-----	1486


```
Db 257398 GTTCTCTCC-AGCAAGTGTGGTCTTCTCGAGGACCTTCCGCCAGCAGGAACTGA 257456
Qy 2323 -----::: :::::|||||
Db 257457 GGAGCGCGGAGTACCGTTGAGCAGAGCGAGACAAACCTCCGGCACTCTCATGGG 257516
Qy 2327 rArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleleAr 2347
Db 257517 TCGGTACGTTCCGAGCAGGAAGAAGAGCTG-----CTCAGCCTCGTCCGCATCCACTC 257570
Qy 2347 GlyAlaLeuMetGlyLysTyrAspGlnTrpGluSerProProLeuSerAlaAsnAl 2367
Db 257571 CGCGGAGTGTCTCGGCGC---GACGACTCCGAGCCATCCCGCGG-TCGGCTGTTC 257626
Qy 2367 aPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAs 2387
Db 257627 GGGATCTAGGTTTCGACTCGTTCGGCGGTGGAGCTTCGCAACCACTCCGACACAGA 257686
Qy 2387 pGly-----ArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLy 2404
Db 257687 CGGAGCTGGCTCTCGCGACGACTCTGCTTCGATTACCCGAGCCCAAGCTCGCC 257746
Qy 2404 sValSerGlyArgProSerSerArgLysAlaLysSerProAla----- 2418
Db 257747 AATTTCGTCT-CCGAGATCGCG---AGTTCAGCGCCGACAACTCAACTCCGCTTCCG 257802
Qy 2419 ---ProGlyLeuAlaSerGlyAspArgProProSerValSer---SerValHisSerGl 2436
Db 257803 CGACCCCGCGGAGCTCGATGAGCCGATCGCCATCGTTGGTCATGGCTGTGCTGCC 257862
Qy 2436 uGlyAspCysAsnArgThr-----ProLeuThrAsnArgVa 2449
Db 257863 GCGGAGTGCACCTCGCGGAGGACTTCTGGGATCTGATCTCTCGAGCAGGACGCGATC 257922
Qy 2449 lTrpGluAspArgProSerSerAlaGly-----SerThrProPheProTy 2464
Db 257923 GCGGATTCCTCCACCGACCGCGGTGGAGCTTGACACGCTCTACGACCCCGACCCGAC 257982
Qy 2464 rAsnProLeu-----::: :::::|||||
Db 257983 CACCCCGGCGCTCTACACCCGAAAGCGGGATTCCTCTACGAGCAGGCGCACTTCGAC 258042
Qy 2468 eMetArgLeuGlnAlaGlyValMetAlaSerProProPro-----ProGlyLeuProAl 2486
Db 258043 GCCGAATTCCTGGGCATCAGCCCCCGGAGACCTCGCCATGGACCCCGACCAAGACTC 258102
Qy 2486 aGlySerGlyProLeuAlaGlyProHisHisAlaTrpAspGluGluProLysPro 2504
Db 258103 CTCTCGAAACCGCTCGGAAACCATCGAACACCGCGGCATCAACCCCGCACCC 258157

RESULT 80
LOCUS AE006968/c 16188 bp DNA linear BCT 27-APR-2001
DEFINITION Mycobacterium tuberculosis CDC1551, section 54 of 280 of the
complete genome.
ACCESSION AE006968 AE0000516
VERSION AE006968.1 GI:13880290
KEYWORDS
SOURCE Mycobacterium tuberculosis CDC1551
ORGANISM Mycobacterium tuberculosis CDC1551
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE 1 (bases 1 to 16188)
AUTHORS Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
Kolony,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
Gill,J., Mikula,A. and Bishai,W.
TITLE Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
JOURNAL Unpublished
```

REFERENCE
AUTHORSTITLE
JOURNALFEATURES
source

```
1..16188
/organism="Mycobacterium tuberculosis CDC1551"
/mol_type="genomic DNA"
/strain="CDC1551"
/db_xref="taxon:83331"
/notes="clinical strain"
111..1436
/gene="MT0756.1"
111..1436
/gene="MT0756.1"
/notes="similar to GP:2911815; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="preprotein translocase SecY subunit"
/protein_id="AAK44990.1"
/db_xref="GI:13880291"
/translations="MISAFISLSRLTDLRRKILFTLGLVILYRVGAALPSGVNFPNV
QQCKEASAGEAGQYSLINLFSGALLKLTFAVGMPYITASIIIVQLLVIPRE
ELRKGAGQSKMTQYTRYIAIALAILQATSIVALAANGGLQCCSDIDAQSIPTL
VTVLVMTGAALVMMGELITRIGNGMSLIFVGAARIPAEGSILBSRGVVF
TAVCAALIIIVGVFVEQQRIPVQYAKRMVGRMYGTTSTYLPKVNQAGVPI
FASLIYPHILITQIRSGVGVGNSWMDKFGTYLSDPSNLVIGIYFGLIIEFTYF
VYSITNPDERADEMKKFGGFIQIRPGRTADYLRVLSRTLPGSIYLGVIAPLN
LPQIAGGCTVQNLPGGTAVLIMIGVGLTVKQIESQLMQRNYEGLK"
1433..1978
/gene="MT0757"
1433..1978
/gene="MT0757"
/notes="similar to PID:2344858; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="adenylate kinase"
/protein_id="AAK44991.1"
/db_xref="GI:13880292"
/translations="MRVLLGPPGAGKGTQAVKLAELGIPQISTGELFRNIEGK
LGVEAKRYLDAGDLVPSDLTDELVDRLNPNDAANGFILDGYPKSVQAKALHEMLR
RGTDIDAVLEFRVSEVLLERLKRGRADDDVILNRMKVYRDETAPLLEYRDLK
TVDAVGTMDEVAFARALALGK"
1981..2781
/gene="MT0758"
1981..2781
/gene="MT0758"
/notes="similar to GB:D00619 SP:Pi19994 PID:216341
PID:1044986 GB:AL009126; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="methionine aminopeptidase"
/protein_id="AAK44992.1"
/db_xref="GI:13880293"
/translations="MRPLARLGRVVPQSRAGELDAMAAAGAVVAALRAIRAAAP
GTSLSLDETAESVIRESGATPSFLGVHGPASTCASINDRVVHGIPTAEVLAPGDL
VSDDCGALVDCWHGDAATITFGVGLSADADALSEATRESLQAGIAMVGNRLDVAAH
AIEGTAAELRYSPFGIVAGVGGHIGQMMDPLPNEGAPRGPGLAAGSLVLA
EPMLTGLTGTIVVLDKMTVTADGSRRAHWEHTVAVTDDGPRLITLG"
2854..3387
/gene="MT0759"
2854..3387
/gene="MT0759"
/notes="similar to GP:5834669; identified by sequence
similarity; putative"
```

```
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma-70 factor, ECF subfamily"
/protein_id="AAK44993.1"
/db_xref="GI:13880294"
/translation="MARVSGAAAEALMRALYDEHAALVLYALRLTGDAQAQADV
QETLRWQHEVIGDTARAKAWLFTVARNMIIIDERSARNRVVSGTDSQGTPEQS
TDEVNALDLRLIALDALQSAEHRRAVITQSYRGWSAQATIDLGIAEGVKSRLH
YAVRALRLTQLQELGVR"
3451..4203
/gene="MT0760"
CDS
3451..4203
/gene="MT0760"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44994.1"
/db_xref="GI:13880295"
/translation="MTMLRGGLGPPDDTGREVSTGDDHHYAMWDAAYVLGALSADR
REFEHLACPECRGAVTELCGPALLSQLDRDEVAIASEAPTIVVASGLSPILLPSL
LAATHRRRLRLITWVASSAAALVLAIVGVQGHSAAPQAPAAVSLPMAQVGTQL
LASTVISGEPWCTFINLRCVCLAPPYASHDTLAVVVGSDGSTRLATWLAEPGHTA
TPAGSISTPVDQIAAQVVAADTGQVLLQSL"
4258..4470
/gene="MT0761"
CDS
4258..4470
/gene="MT0761"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44995.1"
/db_xref="GI:13880296"
/translation="MVGRPPPARSLHDAHLQSSITVVLAKEHFAAATVADGYSRSG
AGFVAAASGGGTFLQKCAATAS"
4518..5015
/gene="MT0762"
CDS
4518..5015
/gene="MT0762"
/notes="similar to GB:AL009126; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, MarR family"
/protein_id="AAK44996.1"
/db_xref="GI:13880297"
/translation="MASDNRDPIAARANWERSGWDVSLGMVAVTSVWRAHOILLAR
VETALRPYLSFSFELLRLIAFSRIGALPITKASDRLOVHTVSTVAIRRLQADGLV
RVPHPTDGRITLQITELGRSTVEDATVTLNEQVFANVGWGAESQALVSAVETLRR
NAGDF"
5012..5224
/gene="MT0762.1"
CDS
5012..5224
/gene="MT0762.1"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44997.1"
/db_xref="GI:13880298"
/translation="WRAGRRVSAOCRAEMGAYASARARLGAQAPATFHDQFVALTRAR
ARMRPPRAGLLNGGARRYPAGS"
5373..5921
/gene="MT0763"
CDS
5373..5921
/gene="MT0763"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44998.1"
/db_xref="GI:13880299"
```

```
/translation="MDPLMAHORAQDAFAALLANVRADQLGGTTPCSSEWTINDLIEHV
VGGEOVQRWAASPIEPARPDPGLVAAHQAAVAHEIFAAPGMSATFKPLPLGEVPG
QVFTGLATDVLT/HAMDLAATQSTDLDELAVERLAARALVGPQFRGPKPFADE
KPCPRPPADQLAAFLGRIVR"
6166..6972
/gene="MT0764"
CDS
6166..6972
/gene="MT0764"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44999.1"
/db_xref="GI:13880300"
/translation="WLTTRAREVALTQHIGVSAETDRAVVVKLRQAYDSILVCGRRRL
CAIGAEINAVAHQALGLDTPAGARNEFRLATKAHDIITRLVLAATAESQAGARLR
SLASYQAVGFGPKPFPDPDPVFPFPYQPKWAACRARGODPKVYVTRTFHAPMSAR
FRSLPAGDSVLVYCGNDKYGLLHIQAKHGRQWHDIDARWPSAGNWRVLYADYAGTALA
YPERVEYNQNDITFAVYRMSLPDGRVYVFTTRVIIISARDGKLIITAFPTTI"
7006..7254
/gene="MT0765"
CDS
7006..7254
/gene="MT0765"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45000.1"
/db_xref="GI:13880301"
/translation="MTGPPRSYTGRRDLIAEKLPYFOISAMLPKNTPTSETABEFW
DNSLWCWSGDRDTGYTRITVTVSIQVADGERAEFGT"
7251..7613
/gene="MT0766"
CDS
7251..7613
/gene="MT0766"

Alignment Scores:
Pred. No.:          Length:      16188
Score:              684.00      Matches:      498
Percent Similarity: 32.80%      Conservative: 162
Best Local Similarity: 24.75%    Mismatches:   802
Query Match:        5.18%       Indels:        559
DB:                  1          Gaps:          85

US-09-522-753-5 (1-2517) x AE006968 (1-16188)
Qy 668 GlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysAlaPro 687
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 16186 CAGCATCATCTCGGTC-GAGCGCGCGAAGCGATGCTGCAACGCCGCGCGGTGCAG 16128
Qy 688 AlaAlaSerGluAlaAlaAlaPheProProValValGluAspGluMetGluAla 707
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 16127 GCCCGCGCGCTACATCAAGCGCTTCAGCCCGTTCTCACCGAGCAGCGCCGCCGCCG 16068
Qy 708 SerGlyValSerGlyAsnGluGluMetValGluAlaGluAlaAlaLeuHisAlaSer 727
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 16067 GCCGCGCGCGCTCGG-----GCCGGGTGCACCCCGGATTTCCCGCGTGTACC 16053
Qy 728 GlyAsnGluValProArgGlyGluCysSer-----GlyProAlaThrValAsn 743
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 16052 -----GCCGGGTGCACCCCGGATTTCCCGCGTGTACC 16011
Qy 744 AsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAla----- 759
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 16010 GCCGTTGCCCGCGTTCGCCCGCTTGCAGATCACACGGCGCTGCCCGCGCGCCCTT 15951
Qy 760 -----LysAspThrGlyGlnAsnGly-----ProLysProProAlaThrLe 773
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15950 GCCCGCGATGTGGAAGACCGGACAAACCGGTGGCGCGCGCGCGGTCACTCC 15891
Qy 773 uGlyAlaAspGlyPro-----ProProGlyProProThrProPro-----ArgArgTh 789
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15890 GAGACCGAAGCGTCCGACTTGCCGCGCGCGCGCGCGCTCCCGCGCGGTGGCGCGCGGT 15831
```


Db 14115 CGCCGTTGCCGATACAGCGGCGCCCGTGGCGCCGACGAATTGCGGTTGATCGGGGCA 14056
QY 1458 --LysTyrAspThrGlyAlaSerThrThrGlySer-LysLysHisAspValArgSerLeu 1476
Db 14055 GCAGCGCGACAGCGCGCGCTCGCGCGCGCATACGCGCTCGCGCGCTCAGGG 13996
QY 1477 IleGlySerProGlyArgThrPheProValHisProLeuAspValMetAlaAspAla 1496
Db 13995 CTTGACGAACTGGCGTAAACGCGCGC-----CCTGTG----- 13960
QY 1497 ArgAlaLeuGluAargAlaCysTyrGluGluSerLeuLysSerArgProGly----- 1513
Db 13959 -----CGTCCCGCCCTGTGATAGGCTGGCTTGGGCAC 13927
QY 1514 ---ThrAlaSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGlu 1532
Db 13926 CGAACACCGCGCAATCGTGTGACACCTCATCGGCACCGCGGTCAACAAATCCGTCG 13867
QY 1533 LeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGly 1552
Db 13866 TGTGACGGCG-----CGCGCG 13849
QY 1553 HisLeuPro-----ArgGlySerProValThrMetArgGluProThrPro 1567
Db 13848 CATTAGCCGAGCTAATCGCGACCGCATCCCGCCAAATCCGTCGCGCGCCACCA 13789
QY 1568 ArgLeuGlnGluGlySerLeuSerSerLysAlaSerGlnAspArgLysLeuThrSer 1587
Db 13788 GCTCGCGCAAAACCATCACCATGACATCGATCGCTCCCAA-----CCCTCACGGCC 13735
QY 1588 -----ThrProArgGluIleAla----- 1593
Db 13734 TGTGCCCGACTACACGCACACACCGTACGGGTAGGAATTAATGCAGAAATCAACGA 13675
QY 1594 -----LysSerProHisSerThrValProGluHisHisProHisProIleSerPro 1610
Db 13674 TATCGCATCCGACCGCATCGAGCGCTTCCACACGTTTCGCAG-----GCGCCA 13624
QY 1611 TyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIlePro--- 1629
Db 13623 GCTGCTGAGCGGCTACGGGTTCGCGCGCGGTTTCCGACAAACGCGCGCGGTTCGCGGC 13564
QY 1630 -----LeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAla 1647
Db 13563 ACAGGGCGCGCGGAACCCGCA-----CCGCGCGATAGCGTTCGACAGCACCC 13510
QY 1648 AlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyr-ProPr 1667
Db 13509 GCG-----ACGGGCTCTGACAAAGCGCTCGAGCGCGGGGATCTTGGCCGAG 13462
QY 1667 oTyrIleuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAs 1687
Db 13461 ATACCTGCGCGCGC-GCCCCCATTTTGGCGGCGCGC----- 13426
QY 1687 nAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAl 1707
Db 13425 -----TCCGAGCGCATCTGCATCACCGCGAGTGGTTGACCGCGGAGCATG 13376
QY 1707 aAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaG 1727
Db 13375 CCGTGCGCGCGGTACGCAACCGTTGAGCCCGCTTTTGGCCCGCAGCAGCAACCGCGCG 13316
QY 1727 Y-Pro-----ArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValP 1745
Db 13315 CGCGCGCGGTTCGCGGCTGCGCGGTGGCGCGGTGCGCGCTGCGCGCTG-----C 13259
QY 1745 roProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlap 1765
Db 13258 CGCGGTGCCGATCAGCGCGCGGTGCGC-----CCGACGCCAC 13220
QY 1765 roGlnProPheSerSerArgHisSerSerProLeuSerProGlyGlyProThrHis 1784
Db 13219 CGGCGCGC-----CCGAAGCCAGCCAGCGGCAATTGCGCGCGCCCGCCACCGACCC 13166

QY 1785 -----LeuThrLysProThr-----ThrThrSerSerSerGlu 1795
Db 13165 CGCGGTGCCGAACCGCGGAACCGCGCGCGCGCGCTCGACAAACAGACCGCGCGT 13106
QY 1796 ArgGluArgAspArgAspArgGluArgAsp-----ArgAspArgGluArgGluLys 1812
Db 13105 TGCCACCGCGCGCGCGTCCACCGAGAAACCAATCCCGCGCGCGCGCGCGCGC 13046
QY 1813 SerIleLeuThrSerThrThrThrValGlu-----HisAlaProIleThrArgProGly 1830
Db 13045 CGCTCCGAACAGCAGCCCTGCGCTACCGCGCGCGCGCGCTCCCGACCGTGTCCAGT 12986
QY 1831 ThrGluGln-SerSerGlySerGlySerSer-----GlyGlyGlyGlyGlySerSe 1848
Db 12985 TTTCGCGCTACCGCGCGGTTCGCGCGCGCGCGCGCGCGCGGAGCATTCGGGATTC 12926
QY 1848 rSerArgProAlaSerHisSerHisAlaHisIleHisSerProIleSerProArgThrG 1868
Db 12925 CGCGCGCGCGCAACCGCGCTGCATTAGTGTGAGCGCGCGGTTCGCGCGTGAC 12866
QY 1868 nAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleTh 1888
Db 12865 CTTCCGCGCGCACCGCATCAGCCGCGC---CAGCAGTCCGCGCGTCCCGCGCGC- 12817
QY 1888 rAlaValGluProSerLysProThrValLeuArgSerThrThrSerSer-ProVala 1908
Db 12816 -----CCACCGCGCGCGCGTCC-----GCCACCAACAGCGCGGTTC 12779
QY 1908 rGProAlaAlaThrPheProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyV 1928
Db 12778 CACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12741
QY 1928 aTyrProThrLeuMetGluProValLeuProLysGluAlaProArgValAlaArgP 1948
Db 12740 -----CCGACCGCGCTCCCGCGCGGTGCCCGCGAGT-----ACGTTGCTGGCTC 12695
QY 1948 roGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerG 1968
Db 12694 CGCGCATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12641
QY 1968 lYLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuValProPro 1988
Db 12640 CC---CCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12584
QY 1988 aSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlas 2008
Db 12583 TACCC---CACAGCCG---GCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12530
QY 2008 erProAspProAlaProAlaProAlaSerAlaSerAspProHisArgGluLysThrGlnS 2028
Db 12529 CGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12489
QY 2028 erLysProPheSerIleGlnGluLeuLeuArgSerLeuGlyTyrHisGlySerSerT 2048
Db 12489 ----- 12489
QY 2048 yrSerProGluGlyValGluProValSerProValSerProSerLeuThrHisAspL 2068
Db 12488 -----CCGGGCGCGCGCGTGGC-----CCGCTGCCAGCGCGCC 12455
QY 2068 ysGlyLeuProLysHisLeu-GluGluLeuAspLysSerHisLeuGluGlyGluLeuArg 2087
Db 12454 CGGGGTGCGC-----TTGCCGATCAGCGGTGCCCGGTGCTGTGGGTGCGCGCGT 12401
QY 2088 ProLysGlnProGlyProValLysLeuGlyGluAlaHisLeuProHisLeuArg 2107
Db 12400 TGATCGCGCGCGCGCGTGT-----GTTTACGGGTGTGACGCGGAGAG-GCGCTGGCGCG 12345
QY 2108 ProLeuProGluSerGlnProSerSerProLeuLeuGlnThrAlaProGlyValLys 2127
Db 12344 GCGTTGAAGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12294

[illegible]


```
QY 2407 yArgProSerSerArgIysAlaIysSerProAlaProGlyLeuAlaSerGlyAspArg-P 2427
Db 77523 -----GGGTCTTACCCGGTCTCGACC 77545
QY 2427 roProSerValSerValHisSerGluGly-----AspCysAsnArgArgT 2443
Db 77546 CTAGCTTATTGGGTGAATAGACTCAGATGGCCCGTCTGCTCGGAATGTCTTACATTGA 77605
QY 2443 hrProLeuThrAnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheP 2463
Db 77606 CGGAATAGGTGGCTAAGTTGGTGTGGTGGCGGCTTAACCGGGAGCATCCCT---- 77661
QY 2463 roTyraAnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProG 2483
Db 77662 -----CCCTCTCTCCACCCTCCCG 77683
QY 2483 lyLeuProAlaGly-----SerGlyProLeuAlaGlyPro----- 2494
Db 77684 GGCTGGACCCGGCGGTGGTGGTGGCGGGGTAGGATTGGTGGCGGTAGGTGTG 77743
QY 2495 --HisHisAlaTrpAspGluProIysProLeuLeuCysSerGlnTyrGlu 2511
Db 77744 CGCGGCATCTCTGGCGCAGT-----TGTTCCCGTTACGAA 77778

RESULT 83
SC0939119/c 299050 bp DNA linear BCT 11-FEB-2003
LOCUS SC0939119
DEFINITION Streptomyces coelicolor A3(2) complete genome; segment 16/29.
ACCESSION AL939119 AL353816 AL356612 AL357432 AL357524 AL358692
AL358818 AL359988 AL391406 AL391541 AL392149 AL392150 AL392178
AL645882
VERSION AL939119.1 GI:24427855
KEYWORDS
SOURCE Streptomyces coelicolor A3(2)
ORGANISM Streptomyces coelicolor A3(2)
REFERENCE 1
AUTHORS Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,
Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,
Harper,D., Batenham,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neill,S.,
Rabinowitz,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrall,B.G.,
Parkhill,J. and Hopwood,D.A.
Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)
JOURNAL Nature 417 (6885), 141-147 (2002)
MEDLINE 21996410
PUBMED 12000953
REFERENCE 2 (bases 1 to 299050)
AUTHORS Bentley,S.D.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: sds@sanger.ac.uk
COMMENT On or before Oct 30, 2002 this sequence version replaced
gi:7242748, gi:7636006, gi:8248766, gi:8249971, gi:8388712,
gi:8439477, gi:8894718, gi:9857143, gi:9886716, gi:10129746,
gi:11595528.
FEATURES
source
1. .299050
/organism="Streptomyces coelicolor A3(2)"
/mol_type="genomic DNA"
/strain="A3(2)"
/db_xref="taxon:100226"
200..1727
/gene="rrnA"
200..1727
/gene="rrnA"
/note="rrnA, 16S ribosomal RNA. Coordinates based on

comparison with rrnB from Streptomyces coelicolor
(EMBL:AL109848)."
2004..5126
/gene="rrnA"
2004..5126
/gene="rrnA"
/note="rrnA, 23S ribosomal RNA. Coordinates based on
comparison with rrnB from Streptomyces coelicolor
(EMBL:AL109848)."
5228..5348
/gene="rrnA"
5228..5348
/gene="rrnA"
/note="rrnA, 5S ribosomal RNA. Coordinates based on
comparison with rrnB from Streptomyces coelicolor
(EMBL:AL109848)."
complement(5412..6080)
/gene="SCO4123"
/note="synonym: SCD72A.09c"
complement(5412..6080)
/gene="SCO4123"
/note="SCO4123", possible two-component system response
regulator, len: 222aa; similar to many eg. TR:O86632
(EMBL:AL031155) possible two-component regulator from
Streptomyces coelicolor (222 aa) fasta scores; opt: 973,
z-score: 1128.5, E(): 0, 69.1% identity in 220 aa overlap.
Contains Pfam match to entry PF00072 response reg.
Response regulator receiver domain, Pfam match to entry
PF00196 GerE, Bacterial regulatory proteins, luxR family
and Prosite match to PS00622 Bacterial regulatory
proteins, luxR family signature."
/codon_start=1
/transl_table=11
/product="putative two-component system response
regulator"
/protein_id="CAB92376.1"
/db_xref="GI:8052435"
/db_xref="GOA:Q9KYR7"
/db_xref="SPTREMBL:Q9KYR7"
translation="MTVLIADDDPLQRFGRMLLESQDDMTVVGAEANGAEARMAA
ELHPDVLMNDVRMPGLDGEATRRIVAGGRSVLLTTFDLDEYAGLKGASGFL
VKDALPEELSLGIRAVATGDVAVPSLTRLRLDAYVHHLPAAGTRDHTPGSPPLRSP
LTEREFTLRVNGQGWNTIEARLHLAESTVKTHVGRILAKTGSRDVQAVTLAYET
KLVG"
complement(5415..5609)
/gene="SCO4123"
/note="Pfam match to entry PF00196 GerE, Bacterial
regulatory proteins, luxR family, score 61.00, E-value
2.6e-14"
complement(5475..5558)
/gene="SCO4123"
/note="PS00622 Bacterial regulatory proteins, luxR family
signature"
complement(5736..6077)
/gene="SCO4123"
/note="SCO4123"
/note="Pfam match to entry PF00072 response reg, Response
regulator receiver domain, score 126.30, E-value 5.7e-34"
complement(6077..7447)
/gene="SCO4124"
/note="synonym: SCD72A.10c"
complement(6077..7447)
/gene="SCO4124"
/note="SCO4124"
/note="SCO4124", possible two-component sensor kinase,
len: 456aa; similar to many eg. TR:O86631 (EMBL:AL031155)
putative two-component sensor kinase from Streptomyces
coelicolor (429 aa) fasta scores; opt: 929, z-score:
995.1, E(): 0, 42.7% identity in 422 aa overlap. Contains
Pfam match to entry PF00512 signal, Histidine kinase and
possible membrane-spanning hydrophobic regions. Contains
possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative two-component sensor kinase"
```


QY	230	IleTyrAspGluasnArgIysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGly	249
Db	123039	-----CCCGATGGAGTCGGGCAGCAGGTCTGCTGCCCTTCACCCGGGGCGCTGGA	122986
QY	250	ProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsn	269
Db	122985	CAACCGTTTCGT-----GATCGCGCGCTCTATAACAGCGCG	122950
QY	270	IleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgAsnHis	289
Db	122949	GGACGTGCCACCAA-----GAGCGACGTGCGCCT	122920
QY	290	AlaArgLysGlnTrpLysGln-----	296
Db	122919	GCAAGCGGCTCAAGAAGAGCCGTGGGCACACCCTGTCCGACCCAGGCCAACCG	122860
QY	297	-----LysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLys	312
Db	122859	GGTGCAGCTCTCAGCGCAGCGCACCGCGGGGCAAGAGGCGGTGCGGATGCGCAGCGG	122800
QY	313	ValGluArgIleGluAsnPro-----ArgArgAlaLysGluSer	327
Db	122799	CAACGACAAGTACCATCAACTTCACCGCCACCAAGACCGAGATCACCGTCGACGAA	122740
QY	328	LysValArgGluTyrTyr-----	333
Db	122739	GGGCTCGTCAGCATCACCGGCAGCGCTCGTGTCGTGGACGGCGGCACCGACTGTC	122680
QY	334	---GluLysGln---PheProGluIleArgLysGlnArgGluLeuGlnGlu-----Arg	349
Db	122679	GCTGAGCGCCAGACGCTCCCTGACCATCAAGAGCGCGGTGCCCTCAGATCCAGGGCGG	122620
QY	350	MetGlnSerArgValGlyGlnArgLysGlyLysSerMetSerAlaAlaArgSerGlu	369
Db	122619	CAGCACGATCAACATCGCAGCGCGGGGCTGTGCGAGTCAACGGCGGAGGCGCCTGAA	122560
QY	370	HisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuLysGlnMet	389
Db	122559	CCTGGGGCGGGCGCGCGCGAC-----GCTGAGCGCGGGCGCGACGACCATCTC	122506
QY	390	ArgGlnLeuAlaValIleProMetLeuTyrAspAlaAspGlnArgIleLysPhe	409
Db	122505	CGCACCGTCAACGTGCAGATCAACTCCGTCATGACGCGGTCTATCGGCGC-----	122455
QY	410	IleAsnMetAengLyLeuMetAlaAspPro-----	419
Db	122454	--GACCTTCGAGT-----CAAGACCCCATTCTTAAGTTCGCGACGATCCCCGTGCC	122404
QY	420	-----MetLysValTyrLysAspArgGlnValMetAsnMet-----	431
Db	122403	GGGCGCTGCAGALCGTTTTCCGAGGAGCGAGGTTC--TCCCTGATGGCCGAACAGTTC	122345
QY	432	-----TrpSer-----	433
Db	122344	GTCGCTCGGCTGCTGCTTCCCTCGCATCGGGCGGACCGCGCGGATCGCCCTCGTC	122285
QY	434	-----GluGlnGlu---LysGluTrpPheArg-----	442
Db	122284	AGCGCGAGCAGAGGAGTGGAGGAGCCATCGCGGTCTATCTCGCCACCGCGCGGGGAG	122225
QY	443	LysPheMetGlnHisProLysAsnPhedly-----LeuIleAlaSerPhe	457
Db	122224	CGGCCGATGGCG-----CCGAGTTTCGGTTCGCCCATCCACGACCTCGTCTTCGGCGCC	122171
QY	458	LeuGluArgLysThrValAla-----GluCysValLeuTyrTyrTyrLeuThrLysLys	475
Db	122170	GTCACAGCAGCAGCGCGCGCATCCAGCAGGAGTGTATGTCACCTTGGACCGCTG-	122112
QY	476	AsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgGlyLysSerGln---	494
Db	122111	GGNACCACGATCGAGGTCCACGAGTCCAGCGTCACCGCGGCGAGGAAACAGAACGTCT	122052

QY	1465	rThrGlySerLysLysHisAspValArgSerLeuIleGlySerProGlyArgThrPhepr	1485
Db	119292	-----AGCGCGGTGCGCGCCGACCG-----CGCGCGCCGCG	119261
QY	1485	oProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGI	1505
Db	119260	GCGGCACGTCCAGCAGCGCCCGGCTTCGACGCTTCGCGGCGCGGCGTGCCTCC	119201
QY	1505	uGlySerLeuLysSerArgPro-GlyThrAlaSerSerSerGlyGlySerIleAlaArgG	1525
Db	119200	GCGGCACCGCGCGCGCCGACCGGACCGCGCGGACCGGACCGCGCGCGCG-----CATCGG	119144
QY	1525	lyAlaProValIleValProGluLeuGly-----	1534
Db	119143	GAGAACCGACGCTGCCCGCCAGCAGC-GGCGTACGCGCGCGCGACCGCGGACCGCGCGG	119085
QY	1535	--LysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisL	1554
Db	119084	AGACGCGCGCGCCACCGCGCGGACACCGCGGACCGGACCGCGCGCGCGCGCGCG	119025
QY	1554	euProArgGlySerProValThr-----MetArgGluProThrProArgLeuGlnG	1571
Db	119024	CGCGCGGAGCGCGCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	118975
QY	1571	luGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgG	1591
Db	118975	-----	118975
QY	1591	luIleAlaLysSerProHisSerThrValProGlu-----HisHisProH	1606
Db	118974	-----GCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	118920
QY	1606	isProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgS	1626
Db	118919	CGCTCCGCGCGCG-----	118906
QY	1626	erHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaA	1646
Db	118905	-----CCCGAACCGGTCCACGACACATGTCGACCGGG-----CCGCTCGCTGCT	118857
QY	1646	laAlaAlaTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrP	1666
Db	118856	CGTCCCGTCCCGACCGCGGAGC-----CCGCGCGCGCGCGCGCGCGCGCGCG	118815
QY	1666	roProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleI	1686
Db	118814	CCCCG-----CCGTGCGCGCGCGTCTCGCGCGCGCGCGCGCGCGCGCGCGCG	118774
QY	1686	leAsnAspTyrIleThrSerGlnGlnMetHisAsnThrAlaThrAlaMetAlaGlnA	1706
Db	118773	-----ACGCGCGCAGGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	118728
QY	1706	rgAlaAspMetLeuArgGlyLeuSerProArgGlySerSerLeuAlaLeuAsnTyrAlaA	1726
Db	118727	CGGCGCGCGCGCGCGGTCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	118668
QY	1726	laGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProp	1746
Db	118667	CTGCGGCA-----TGCCCTCTGTCTCGAGCGCGCGCGCGCGCGCGCGCGCGCG	118617
QY	1746	roThrProGly-----	1751
Db	118616	CTGTGCGCGCGGTCTTCGCGGACGCGGTTCAGAGACCGCGCGCGCGCGCGCGCGCG	118557
QY	1751	roAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerA	1771
Db	118556	CGGCTGCGCGCGGTCTTCGACCGCATCGGACCTGGATCATCCGTCGCGCGCGCGCG	118497
QY	1771	rgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrT	1791
Db	118496	CCTGCGCGCTCTCG-----GCTTCATGATACATCCCGACCGCGGTCCAGAACACCGCGG	118440
QY	1791	hrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgG	1811
Db	118439	CTTCCCGAGCGCGCGCGCGCTCCCGGACGGCATCAAGGCTCGC-----	118393
QY	1811	luLysSerIleLeuThrSerThrThrValGluHisAlaProIleThrPargProGlyT	1831
Db	118392	-----GCTCCT-----TCCCGCGGCA 118377	
QY	1831	hrGluGlnSerSerGlySerSerGlyGlyGlyGlyGlySerSerSerSerArg- 1850	
Db	118376	CAATCCCGAGCTCTCATCGACAGCTCAGCAACACCTGGTGGGGCGCGGCATCACGGA	118317
QY	1851	--ProAlaSerHisSerHisAlaHisGlnHisSerProIleSerPro-----ArgThrG	1868
Db	118316	GTCCGCGCAGCGCCAGTGGTCTGAGCGAGCTTACCGAGCGCGACCGGCTCTCGACGT	118257
QY	1868	InAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleT	1888
Db	118256	GATCATCACCGCGCGGTCTCCACCGCGCGCGCGAGCTGCGGAGTCCGCTCGCGCA	118197
QY	1888	hrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSerSerProValA	1908
Db	118196	CCGCG-----TCAGCGCAGCATCACCATGAAGCGGTCGGTCTCA 118155	
QY	1908	rgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyV	1928
Db	118154	GACCGCGACCTCA-----CCTGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG	118101
QY	1928	alTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgP	1948
Db	118100	CG-----TCGCGCAGGTCCACCAAGTCCGCTTCAC 118071	
QY	1948	rgGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaAlaArgS	1968
Db	118070	CATCGAGTCCGCCAG-----CGCGCTCCCGCAAGAGCAGGT 118032	
QY	1968	lyLeuGluProLaserSerProSerLysGlySerGluProArgProLeuValProProv	1988
Db	118031	GGCATCTCCGAGATCGAGTCTCTTCG-----GCCGTCCAGCGCGCAACCGCTCTCTG	117978
QY	1988	alSerGlyHisAlaThrIleAla-----ArgThrProAlaLysAsnLeuA 2003	
Db	117977	ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	117918
QY	2003	la-----ProHisHisAlaSerProAspProProAlaProProAlaSerA 2018	
Db	117917	CCCCACCGACCGGACCGCACCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	117861
QY	2018	laSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuLul	2038
Db	117860	CCATGCCCGCGCACGAGC-GATATCCACCGCGCGTCT----- 117824	
QY	2038	euArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerP	2058
Db	117823	-----GCTGTCCGCGCGCACCTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT	117769
QY	2058	roValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluLeuA	2078
Db	117768	CGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	117736
QY	2078	sPlySerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyG	2098
Db	117735	ACCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	117677
QY	2098	lyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerP	2118
Db	117676	-----CTGCCC-----CGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	117637
QY	2118	roLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValThrLeuAlaGlnH	2138
Db	117636	CC-----CGCGCGCTCTCC-----CGACGGCACCGCGCGCGCGCGCGCGCG	117589
QY	2138	is-----IleSerGluValIleThrGlnAspTyr-----ThrArgHisHisP 2152	

```
Db 117588 ACGCCGTTGGACTGGAAGAACCCCTCCGGCGAGACCATGGACCTCGCACTCGTCGGGCC 117529
Qy 2152 roGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysp 2172
Db 117528 GGCACGAGCGCACCCCGGACGACCGGATCGGCTC-----C 117493
Qy 2172 roValLeuAspLeuArgPro-----ProSerAspLeuTyrL 2185
Db 117492 CTGCTCTTCAACTTCGGCGCGCGCGCGGCTCCGGCGTGGCCACCTTCGG- GCCTGGC 117434
Qy 2185 euProProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGly----- 2202
Db 117433 CGCCGACTACGAGAACTCCGACACCCGCTACGACTCGCTACGCCCGCGCGCGT 117374
Qy 2203 -----GlyLysArgSerPro---GluProAsnLysThrSerValL 2215
Db 117373 CGGCAACAGCGCGCGTGGCTCCCGGAGCGCGCTCCCGACGAGGACGAGA----- 117318
Qy 2215 euGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProG 2235
Db 117317 -----TCGACGGCATCCCGGACGACGCGGGCGGACGAGACCGA 117281
Qy 2235 lyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProS 2255
Db 117280 CGCGCTCGCTCGCTCAACC-----GGGAGACCGCGCGCGCTG 117242
Qy 2255 exArgMetGlySerLysSerProGlyAsnThrSerGlnProPro----- 2269
Db 117241 CGAGAGAACTCCGGCAGCGCCCTCCCTAGTCGGACACCCAGCGCGCGGACAT 117182
Qy 2270 -----AlaPheSerLysLeuThrGluSer-----AsnSerAlaMetValLysS 2285
Db 117181 GGACCTGATCGCCAGGTCTCTCGCGACGAGAACTGCATCTCGCATCTCTCTACGG 117122
Qy 2285 erLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrA 2305
Db 117121 CACCGAACTCGCGCGCTCTACGGCGCATTTGTTCCCGGAGCAGCTCGCGCGCGTACT 117062
Qy 2305 snlle-----SerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyT 2322
Db 117061 CGACGGTGTGTCGACCCGACCCAGGACAGA----- 117030
Qy 2322 hrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyL 2342
Db 117029 -----TGAGGAGAGCCCTCGCCAGCGCGCGGCTTCAGCTCGCCT---- 116988
Qy 2342 euGluAlaIlelleArgLysAlaLeuMetGlyLysTyrAspGlnTyrGluGluSerProp 2362
Db 116987 --TCGAGCACTTCGCCGCTGTGCGCGCGGAGGCTGCACACTGGCGGAGGACGCGA 116930
Qy 2362 roLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaMetP 2382
Db 116929 GGCATCGTGGACCTCGCTCGCGCTCGGAGCGCGCTCGGACGACACCCCTGTCCAC 116870
Qy 2382 rolleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyL 2402
Db 116869 CCGGACGACGCGGAATCG----- 116850
Qy 2402 ysaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuA 2422
Db 116849 -----ACGCGACGCGCTCGTTCGCGCGCGTCCACCGCGCGCTCTACCGGACGGA 116801
Qy 2422 laSerGlyAspArg-----ProPro-----SerValSerS 2432
Db 116800 CTATGCGCGCGCTCGGCTCGGTCGTCGGAAGCGCGCGCGGACGACGACGCGCGCTCT 116741
Qy 2432 erValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluA 2452
Db 116740 GCAGGACCTCCGGAAGCCCTCGCGACGACGCGCACCGTGGCGGACGAGCG-----ACGG 116684
Qy 2452 spArgProSerAlaGlySerThrProPheProTyrAsnProLeu-1leMetArgLeu 2471
```

VOTVI PARTAWLDRHGVRAWEPAAREROTDRKHSCKDKGKACGCEGIVALLAWFEBAE
 KLLRELGDVFTGIAQEARASAGVTLVWSQKASGYQLSTDTRASLPAAMCFGKGDGA
 SFALPEVLDADPAAGNKRKYVYLASAGVDEDLHATPVKTYWTGSGBYERNWAE
 YVVTQETTRAAIDAVTADAATKFAKFFTRDRARAFAGAPAGARILLTAAODVDQVE
 KELVDADADVDPTAIDILRPFTPLPPSPASKPGPEAREBELDELVRMLAGMGFGTVAV
 HDLGPVLEOLGRDRSWSVKEMKRLAEBGLVPVLAVY"

gene
 3988. .4500

/gene="spdB1SL"

CDS
 3988. .4500

/gene="spdB1SL"

/codon_start=1

/transl_table=11

/product="putative spread protein"

/protein_id="BAC21267.1"

/db_xref="GI:23978904"

/translation="MTDRPQYPTGRQITDADLAAALLCPTRPAPPRPPAARPGPG
 PAVDVAALLEFAROGALEAAVTQTAPAAETSASILPRWAVGTAVASVIGIAPCLLG
 WADLLATGAAVAAGESPADPQCILIGAVLDRQPSIARSAPKGGGIEVTQITQTI
 TQSVKGGGR"

gene
 4497. .5723

/gene="spdB2SL"

CDS
 4497. .5723

/gene="spdB2SL"

/codon_start=1

/transl_table=11

/product="putative spread protein"

/protein_id="BAC21268.1"

/db_xref="GI:23978905"

/translation="MKLDTHAVRSTGRSRLRLVWLIVFGAVLFSVLTVTPLVERVTP
 EGMGTAWILPFVVDAAVTVVRVDEITRLDRAGWGWLALRWLTGWTALNIGDS
 MLKGDVGVGHVAVPTLLIFTAEALGWERAJTRAVARIESRAERAGARALSART
 VSAQSAQRDARAAREQORLADAREQORODREHAREQOQALRELRANDAEQAR
 AEREHAQVERERREDAOREEQORADAKADAEAREQOKEATRELRANDAEQAR
 AEARRAPVSAVSTPRPTVSTVTPAHDITTKTSETPKAAEARRPATPVSAVST
 PRPTVSTVTPAHDITTKTPAETLKMTEADARRAAVEAREGRSQRQVAALTSVGVW
 VAKRFDOLEOTGVSA"

gene
 5720. 6166

/gene="spdB3SL"

CDS
 5720. 6166

/gene="spdB3SL"

/codon_start=1

/transl_table=11

/product="putative spread protein"

/protein_id="BAC21269.1"

/db_xref="GI:23978906"

/translation="MSTDDLNAHAEVRAEITRTDTKAGLLIAFVGAVIAGAGSIAHD
 VPMTLPAYIAGGIGCAALLAAAVLLHVRPLGRGHGWPLWATLTPQQLAAEPRP
 GRDVVGLSRLLAAAKFTGLRQAILDLITAGAAVFVAAVALLALGAA"

gene
 8113. .8865

/gene="orf753"

CDS
 8113. .8865

/gene="orf753"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAC21270.1"

/db_xref="GI:23978907"

/translation="MKWEDRATTDREKIRAGAGQGGPNVGLATPGSLGVVDLMPKE
 GASLGAIEVAFALCERAGQPLPATFRAGTASGGEHLVFTAPVGIRLGNSAGLLGLV
 DTRWAGGVVAPGSIITPGVYVTLDDAPVLPPLPSWLLLEALTAROKSRRPRGGVPAPFA
 VSGSRAARALEREQGNVQAAPTQKANTLNRSALVGRFVANGDIDRQVVEBAFQGA
 GERGLTAACRATIRSLDSSLKARPREVA"

gene
 9198. .10739

/gene="orf1542"

CDS
 9198. .10739

/gene="orf1542"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAC21271.1"

/db_xref="GI:23978908"

/translation="MTAHPDELWAGFDENSAEDVAGPIWDEPVPLNPRGQLPVFPVDA
 LPDLWAAMTAGVAEETQTPVDLAGCLALAVVGTAAGRLTVTVRGQWSEBPVNLVATA

LPGNKRKSAVFGMLTNPLLAERKALIELTAPORAEAAASARIKAAAEKAEKLAANAE
 ADKLADLTAQAVSLSEAAERAVI PAEPOLVADVDTAESLTTLLAQODGRISILSPRGE
 IPEILTAGRYSGAPNMGIELKHAGDMARVNRQADQAYTENPAITWGLAVQPDVLDAI
 GQIKAGDGGGLLARTLYSRPELSLVGRNLTPELLSPDIAITYAKLGRLLALHLAGWTE
 TAEFLTTPADADLILAYQRVTSRSLRKGDLAPIINWASKRQGVARIAGLLHLHAGHP
 EDAMWTRPTAADTMAATRLGDYTFTHALDFVFNAMKADPAQAEARTLLTLTETRTAAF
 TKRDLFRAMPSEFPTMSDLDPALNLLHEHGVWRQPPPTTRTGRGPPSPRYETHPR
 ITPAA"

gene
 11232. .12608

/gene="int"

CDS
 11232. .12608

/gene="int"

/codon_start=1

/transl_table=11

/product="putative integrase"

/protein_id="BAC21272.1"

/db_xref="GI:23978909"

/translation="MGRATGTQRRRLRHAPARAARSSGAERAPFYSERIHSTWAGH
 VODRYKTELGSDDGKPRVKTRFGTGRYARYVPGDTEKSKSPDRQKRLAEQWL
 NETAADMARGHYIDPRAARITPRQAEKWVATHTEVNSKEAERLRHLHVYVIGTR
 PLASQOPGHRTWLGELSKYPAASHRRIVFGTVSAAGALDDGLIAKNPKKARSIO
 IPKPIPTRVPTWTRQVFGVRAALPERLRATVDVAGCGLRQGEVGLSEDELDPRGG
 WLSVNHQLKRIIRGKFVFLPKGAKVRDVLPTDVALQEHQRKQYPSVAVTLPMWTPD
 GLVTKSLIFSGSTRDHVRVSHFNHMKPALAAGVILPELVEGKRSPSEHGMHALR
 HFVASVLLDAGENIRALSOYLGHADPGFTLRTVTYTHLMPSSSEGRTRKAIESLYQQLRDH
 DHGPETAQ"

gene
 12990. .13451

/gene="korSL"

CDS
 12990. .13451

/gene="korSL"

/codon_start=1

/transl_table=11

/product="putative transcriptional regulator protein"

/protein_id="BAC21273.1"

/db_xref="GI:23978910"

/translation="MPQASPRGTYLKVADAKVAKQIDAEAAAMTELPFLAELMAAYGV
 RGVAVRAPVKVKEEGRABEVRGWRVVRDGEQDRRPLVERIADYFRVDGLKVGDDF
 PSTSKLADRFGASRPTVAKALDRLETSGLSASRQGRKRTVLALPQREEQS"

gene
 complement (14012. .14485)

CDS
 complement (14012. .14485)

/gene="mutTSL"

/gene="mutTSL"

/codon_start=1

/transl_table=11

/product="mutT-like protein"

/protein_id="BAC21274.1"

/db_xref="GI:23978911"

/translation="MGRIDYLDHDPDAPPANSVVSVVAFVQDQGRVLMIQRSNDGRW
 ALPGGGHDAGESISDVTVREVWEETGIDAEVIDSGIYTDPGHVMAVDGGEIRQQFSI
 CFRARPTGGEVRTSETTQVRWVAPADLVDELVHPTWRLRIEHMDRTRTAPYIG"

ORIGIN

Alignment Scores:

Pred. No.:	1,566-05	Length:	15397
Score:	673.50	Matches:	652
Percent Similarity:	31.94%	Conservative:	295
Best Local Similarity:	21.99%	Mismatches:	1156
Query Match:	5.10%	Indels:	872
DB:	1	Gaps:	122

US-09-522-753-5 (1-2517) x AB093554 (1-15397)

Qy	45	HISerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnArg	64
Db	3815	CACAGCGGCACATCCCGCCGACAC-----CCGACAGCGCGGTATCCCCACATCAGCGC	3868
Qy	65	ArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnLeuHisLeu	84
Db	3869	GGGTGTGAATTAGGGCGGCACACCGGCCCGCCCAACACCCCGCGCTACACCCCGCACACC	3928
Qy	85	ArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGlu	104
Db	3929	AAACACGAGAGCGCGACCCCGGTTGCCCGC-----GATCGG	3964

Qy 1313 tMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyAr 1333
Db 7932 ----- 7932
Qy 1333 gAlaIleProGluArgHisSerProHisHisLeuLysGluGlnHisIleArgG1 1353
Db 7933 ----- ACCAGCATGA-CCCAACCGA-----CCCTC 7957
Qy 1353 ySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgAr 1373
Db 7958 ATCCGGCAGAGCGGCCCCCGC----- 7982
Qy 1373 gGluAlaLysLeuLeuLysArgGluGlyThrProProProProProProSerArgAspLe 1393
Db 7983 -----CGACGCTGTCGCCGCCCTCCAGCGCCCGCGAGCGCGCTGG 8023
Qy 1393 uThrGluAlaTyrLysThrGlnAlaLeuGlyProLeu-----LysLeuLysPr 1409
Db 8024 CACGTCTTCCCACTCCCGCCGGGTACGAAGCGGCTCCCTGCACGCGGAGCGCTGCC 8083
Qy 1409 oAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIlePr 1429
Db 8084 CCGCACCGCCACCTCGGTGACGACGACGACGACGACGACGACGACGACGACGACGAC 8143
Qy 1429 oArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluG1 1449
Db 8144 GCGAGCGGAT-----CGCGG 8158
Qy 1449 ySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLy 1469
Db 8159 GCGGTGGGGCGAGG---CCCGTTCAACGTCGCTTGGCCACGCGGCGCTCGGGGCTCG 8215
Qy 1469 sLysHisAspValArgSerLeuIleGlySer-----ProGlyArgThrPheProPr 1486
Db 8216 TCGTGTGACCTCGACATGCCCAAGAGGCGCCAGCGGCTCAGCGGTGCGAGGTGT 8275
Qy 1486 oValHisProLeu-----AspValMetAlaAspAlaArgAlaLeuGluAr 1501
Db 8276 TCGCGGCCCTCTGTAGCGCGCGCGCCAGCGCTTCCGCGGACCTTCCGGCGAGGACG 8335
Qy 1501 gAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlyse 1521
Db 8336 CGTGGGTGGCGAGCACCTCTACTTACCGCCCGGTCG-GCATCCGCTCGGCACTCC 8394
Qy 1521 rIle-----AlaArgGlyAla-----Pr 1527
Db 8395 GCTGGATTGTAGCCAGTTGGTGGACACGCGGCGGTGGGGCGGCTACGTCTGCGCACCC 8454
Qy 1527 oValIleValProGluLeuGly----- 1534
Db 8455 GGGAGCATCACCCCGCGGGGTGTACACGGTCTCTGGACGACGCTCCGCTCTCTCTG 8514
Qy 1535 -----LysProArgGlnSerProLeuThrTyrGluAspHi 1546
Db 8515 CCCAGCTGGCTCTTTGAGGCACTCACGCGCGCGCAGAAAGTCT-----CA 8559
Qy 1546 sGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProTh 1566
Db 8560 CGGC---CCCGTGAGGGGTTCGGGCTTCCCGCGGTCTAGTGGCAGCCGCGCGGCGCT 8616
Qy 1566 rProArgLeuGlnGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuTh 1586
Db 8617 GCGGCGC-----TGGAGCGGAGTCCGCGCAACGTCACGCGCGCGCGCGCAGC 8661
Qy 1586 rSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHisProHi 1606
Db 8662 AAGCAGGCCAACACACGCTCAACCGGTCCGCATTCCTG-GTGGG----- 8706
Qy 1606 sProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSe 1626
Db 8707 -----CGTTCGTGTCATGGGGCGACATCGACCGCA 8738

Qy 1626 rHisIleProLeuAlaPheAspProThrSerIleProArgGlyIle-ProLeuAspAlaA 1646
Db 8739 GGTGTGGAGGAGCATTCAGGGGGGGGAATCGGGGACTCACCGCGCGCGAGTG 8798
Qy 1646 laalaIaTyrTyrLeuProArgHisLeuAla-----ProA 1658
Db 8799 CCGCG-----CCACGATCGGTAGCGCCTCGACAGCTCCCTCCCAAGGCCG 8846
Qy 1658 snProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaA 1678
Db 8847 GCCCGGGAGGTGGCATGAGCCCGCAGAAATCCCTCCCTTGGATCGCTCC----- 8899
Qy 1678 laLeuGluAenArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMethHisAla 1698
Db 8900 -----CACCAAC 8906
Qy 1698 snThrAla-----ThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerP 1715
Db 8907 ACCCGCGCTCGCGCGCGAGCGGTAGGACTGCGGGCGGTGGCGCGCGCGAAAGGTC 8966
Qy 1715 roArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGly-----I 1731
Db 8967 GTCCCGAAGCGCTCTTCGGGTCTGACCCGACCCCGCGCGCGGTGAGGATCTCTCGC 9026
Qy 1731 leIleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrP 1751
Db 9027 GATGTCCACATCGGCGCGCGCGCAACGCGATGCCACGCGCGCTCTCTGTGCTC 9086
Qy 1751 roAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerA 1771
Db 9087 CTGCGCGCGGACCTGAGCGCTTCCGCCACACGACAGATCGCCC----- 9130
Qy 1771 rGHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrT 1791
Db 9131 -----GCCTCGACCGCACACCCCGC-----ACCGCGCCAGTGCCTC 9170
Qy 1791 hrSerSerSerGluArgGluArgAspArg----- 1802
Db 9171 GCTGTCAACCAAGGAAGAGCGCATGACCGCGCACGACCCCGAGCTGTGGCGCGG 9230
Qy 1803 --GluArgAspArgAspArgGluArgGlyLysSerIleLeuThrSerThrThrValG 1822
Db 9231 TTCGACGAGATGACCGCGGAGGATGCGCGCGGCCCATCTCTGGGACGACCGCTACCCTC 9290
Qy 1822 luHisAlaProIleTyrArgProGlyThrGluGlnSerSerGlySer----- 1837
Db 9291 AACCGCGCGGCA---GCTCCCGGTGTTCCGTTGACGCGCTGCCGCTGCTCGCC 9347
Qy 1838 -----SerGlySerSerGlyGlyGly-----GlySerSerSerA 1850
Db 9348 GCGATGACGCGCGGGTGGCGGAGGAGACCCAGACCCCGCTCGACCTGGCGGGGTCTC 9407
Qy 1850 rgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspA 1870
Db 9408 GCCTCTCGCGCTCGCGACACCGCGCGCGGTGCGCTCACCGTACCGCTGCGCGG---G 9464
Qy 1870 laLeuGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlav 1890
Db 9465 CAGTGAGCGAGCGGTCAACCTCTACACCGCGCT----- 9499
Qy 1890 alGluProSerLysProThrValLeuArgSerThrSerThrSerSer---ProValArgP 1909
Db 9500 --CGCGC-TCCCGCGCGCAACCGCAAGAGCGGTCTTCGGGTCTATGACCAATCGCT 9556
Qy 1909 roAlaIaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValT 1929
Db 9557 CCTCG-----CCGCCGAGAGCGCTCATCGAACTCACCGCGCGCGCGCGCGCA 9607
Qy 1929 yrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProG 1949
Db 9608 GCGCGCGG-----CCTCGCGCGGATCGCGAAGCGCGCGCGCGCA 9646
Qy 1949 luArgProArg----- 1952

Db 55458 GTCCGCTCCACAGCATCCACATCGGACACGACCAACCCCGCACTCGTAGTGTGGT 55399
 Qy 515 --SerSerGlnGluGluLysAspGluLysGluLysGluLysGluLysGluGluG 534
 Db 55398 GATCACCTTTGTGGAGGACCATTCGCGCGGTCAGTCGTTGAGCGCCCTCTG 55339
 Qy 534 luySProGluValGluAenAspLysGluAenAspLysGluAenAspLysGluLysThrAspThr 554
 Db 55338 ATTCACCGCCAGCCAGAA-----CCACCGCAACACCCGAT--G 55300
 Qy 554 erGlyGluAspAsnAspGluLysGluAValAlaSerLysGlyArgLysThrAlaAsn 574
 Db 55299 CCCATTACGAGCGGATCCACAACTCTCCAAAGCACTAGCCAGCACTCCGCCCA 55240
 Qy 574 erGlnGlyArg--ArgLysGlyArgLysThrArgSerMetAlaAsnGluAlaAsnSerG 593
 Db 55239 GCCGTCCTCGCGCGCTCCGCGAAGACTTGCAGCGCCCATCCGTTGCCAGCCCG 55180
 Qy 593 luGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSer 613
 Db 55179 CTGACGGAGNACTCCAAACAGTCCGCG-----55151
 Qy 613 erArgTrpThrGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgA 633
 Db 55150 -----G 55150
 Qy 633 snTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheT 653
 Db 55149 GGTGCCCATCAGCTCAGC-----55130
 Qy 653 yrPheAsnTyrlsLysArgGlnAsnLeuAspGluLleLeuGlnGlnHisLysLeuLysM 673
 Db 55129 -----CACCGCAAGGCGAGAT-----55112
 Qy 673 etGluLysGluArgAsnAlaArgLysLysLysLysAlaProAlaAlaAsnSerGluG 693
 Db 55111 --CACACTACCGGAACAGTGTGATGACCTGCAGGTGCAACCCAGCGATGACGA 55054
 Qy 693 luAlaAlaPheProValValGluAspGluMetGluAlaSerGlyValSerGlyA 713
 Db 55053 ACACCGCGTGTCCACCGTCACCGCAGGACCTCGAAACCAAGCAATACCGCACCCGCGC- 54995
 Qy 713 snGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProA 733
 Db 54994 -----CGA 54991
 Qy 733 rgGlyGluCysSer-----GlyP 739
 Db 54990 GGCCACACTGCCCGCTCCATTACCGAGATAGCTTCGAAGCCCTCCGCACTCCGCGC 54931
 Qy 739 roAlaThrValAsnAsnSerAspThr-----GluSerIleProSerProHist 756
 Db 54930 CGCGAACCGTCCGCGGTAGTCGTGTATCATCAGCCCGCGGAACACACCCGTCGCGCTCC 54871
 Qy 756 hrCluAlaAlaLysAspThrGlyGlnAsnGlyProLysProAlaThrLeuGly---- 774
 Db 54870 GCGCAGCGACCCCGGATCGATCCAGCGGTTCAAGCGCTCCAGGACAGTCCAGCAG 54811
 Qy 775 -----AlaAspGlyProProGlyPro-----782
 Db 54810 CAACCGTTGTCGGGTTCATCCGAGTGCCTCAGCGAGAAATCCCGAANAACCGCGC 54751
 Qy 783 --ProThrProArgArgThrSerArgAlaPro-----793
 Db 54750 ATCGAACTCCGCGCGTCCCGAGGAACCGCCCTCAGCAGATAGTCCGGAACCG 54691
 Qy 794 -----IleGluProThrProAlaSerGluAlaThrGlyAlaProT 807
 Db 54690 ATCGGATCTGGATCGAACAATCCGCGAGATCCAGCGCGATCCGCGAGTCCGGAAGTCCG 54631
 Qy 807 hrProProAlaProProSerProSerAlaProProProValProLysGluGluL 827
 Db 54630 GACCGCATCCACGCGCCCGCGCAACCAAGCTCTCTCCGGTGAGGACACACCCACC 54571

Qy 827 ysGluGluGluThrAlaAla-----833
 Db 54570 CGGATACGACACCCATCCGAGCTATCGGGACGAGATCGTCTTCGACGCTCGCGTCTG 54511
 Qy 834 --AlaProProValGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 553
 Db 54510 GGTCTCACCGCGCGCGCTCGTCCGACCTCGCTCCAGTT-----54467
 Qy 853 alaSPThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGluG 873
 Db 54466 -----CCTGTGTGAGGAACCGCGCCACGCGCGGTGCGGTAGTTCGTAATACGCGGT 54412
 Qy 873 lyProAlaLysGly-----LysAspAlaGluAlaAlaGluAlaThrAlaGluG 889
 Db 54411 TCCCGGAGCGCGGCTCCAGCAGCCCGCGCGGATTCAGCAAC--TCCACACACGACA 54353
 Qy 889 lyAlaLeuLysAlaGluLysGluGlySerGlyArgAlaThrAlaLysSerS 909
 Db 54352 ACGAATCAAGCCCGA-----TCCGGAAGACGCTCGACCCCGACCT 54308
 Qy 909 erGlyAlaProGlnAspSerSerAlaThrCysSerAlaAspGluValAspGluA 929
 Db 54307 CGGCGCACCCGAAATACCCAGCACCTCGCGGTGCTCCGACACCCACCCCAAGCAAG 54248
 Qy 929 laGluGlyGlyAspLysAsn-----ArgLeuLeu-----Ser-Pro 940
 Db 54247 CGTCTCTCGCGAGGTGCGTCTGCTGCTGCTGATAGAGTGTAGTAGTACCA 54188
 Qy 941 ArgProSer-----LeuLeuThrProThrGlyAspProArgAlaAsn-----954
 Db 54187 GTGCGGTGAGCATCCGCGGTAGCGTTCCGCGCGGTGCTCGCGAGGAGGCCCTGT 54128
 Qy 955 ---AlaSerProGlnLysProLeu---AspLeuLysGlnLysGlnArgAlaAla 972
 Db 54127 CCAACTCTCCCGAACCACCGCTGCGCGATCCACGCGCGCGCATCGAACAGCCCA 54068
 Qy 973 ileProIleGlnValThr-----LysValHisGlu 983
 Db 54067 AACCATCTTGTGTGACACACGCTGAGTGGAGCGCGATCTGCTCGGTGCGATCAC 54008
 Qy 984 ProProArgGluAspAlaAlaProThrLysProAlaProProAlaProProProGln 1003
 Db 54007 TGAGCGCGCGGTATCCCGCTCGACTGCTCCACAGCCCGCCACGCAACGACGCGAG 53948
 Qy 1004 AsnLeu---GlnProGluSerAspAlaProGlnLysProGlySer-----1017
 Db 53947 GCAATCCCTGCGCGCGGTATGCGCAATGCTTCAGGAACGTTTTCGCGCGCAT 53888
 Qy 1018 -----SerProArgGlyLysSerArg-----1024
 Db 53887 AATTACCTGACCGCGCGCGGTCCCGCGACGAGCGAGGAGAACACGCGAAGAACG 53828
 Qy 1025 -----SerProAlaProAlaAspLys 1032
 Db 53827 ACAAGTCCATATCCCGAGTCAGCTCGTGCAGGTTCACGCGCGCGCTTCGGAACGA 53768
 Qy 1033 GluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProCysTrpThrSerGly 1052
 Db 53767 ACACCG-----CCCCCAGCTCTCCGTGTGACGACTCGA 53732
 Qy 1053 LeuPro---PheProValProProArgGluValIleLysAlaSerProHisAlaProAsp 1071
 Db 53731 TCACCCGTGACGCGACACACCGCGG-----CATGACACACAC 53693
 Qy 1072 ProSerAlaPheSerTyAlaProProGlyHisPro-----LeuProLeuGlyLeuHis 1089
 Db 53692 CACACAGC-----GAAACTCCGCGCGGACCGCGCGCAACCCCTCCAGACGACAC 53639
 Qy 1090 AspThrAlaArgProValLeuProArgProProThrIleSerAsnPro-----Pro 1106
 Db 53638 GATCCGCAACGTCCGACGCGCCACATCGACCCCGCACCGCAACCCCTCCAGTCTACCCA 53579

Db 51702 GCCGACCCCTGCTGCGGAGCGCCGGAACCCCGTGGCGGAAGCGAGCGGCTCACGCAC 51643
 Qy 1684 -----ThrIleAAspTyrIleThrSerGlnGlnMetHisHisAsnT 1699
 Db 51642 CTGACGCACCCAAATACCTGGTGTAGCGAGCATGCCAGCGGTGTGAGCTCACCGCTCAA 51583
 Qy 1699 hrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerS 1719
 Db 51582 CGTCGACACGACCGCGAGGTGCG-----AACCGCGGTAGTCACGCG----- 51542
 Qy 1719 erLeuAlaLeuAsnTyr-----AlaAlaGlyProArgGlyIleIleAspL 1734
 Db 51541 -----TGCCTGGCATATCTGTGAACCTCTGCAGCATCGGTCCATCCGATACGAATGAAA 51487
 Qy 1734 eu-----LeuValProProThr-----SerGlnV 1737
 Db 51486 CGCATGCGAAACCGCAACACCGCCACCGAACCCCTTGCCCATCAGCCCGCGCGCAAT 51427
 Qy 1737 alProHisLeuProVal-----LeuValProProThr----- 1747
 Db 51426 GTCTCAAGCACCTCCCGATCGCGGAGAGACACCGACTCAGAAGCGTTGACCGCGGC 51367
 Qy 1747 ----- 1747
 Db 51366 GATCCCCACCCCTTACCCCGAGCGGCCAACACGCGCGAGCTGCTCTCGCCAGCAGC 51307
 Qy 1748 -----ProGlyThrProAlaThrAla----- 1754
 Db 51306 CACCGCAACATGCGACCGGCATCCGGCAACGGCTGCATCAACCGGGCAGCACCCGCCAC 51247
 Qy 1755 -----MetAspArgLeuAlaTyrL 1761
 Db 51246 CAACCGAGCGCGATCCGGGAACAGACAAACACCCCGCGCAACCGCGCCGCAAACTCAC 51187
 Qy 1761 euProThrAlaProGlnProHisSerSerArgHisSerSerProLeuSer----- 1778
 Db 51186 CACCGAATGCCCCATCACACACAGCGCCGAAACACCCACAGCGCCAGCGCTCAGCAA 51127
 Qy 1779 -----ProGlyGlyProThrHisLeuThrIlySerProThrThrThrSerSerGluArg 1797
 Db 51126 ACCAACCTGCAACGCGAACAACCCGACTCGGCCCAACAGTCCGATCCAGCAAGGACCC 51067
 Qy 1797 luArgAspArg----- 1800
 Db 51066 GTCGGAACCGAACAACACATCCCGAACCCCACTGCTGTCGCCAGATCGCATCCAGTCC 51007
 Qy 1801 -----AspArgGluArg-AspArgAspArgGluArgGluIlySerIle 1814
 Db 51006 GGCACAAGCTTCGTCAAAACGCTCAGCGAAACCGGAAACCGCGAGTAGAGCCCTCTTCC 50947
 Qy 1815 -----LeuThrSerThrThrThrValGluHisAlaProIleTyrArgProGly 1830
 Db 50946 CATCCCGAGCCACTGACC---ACCCTGACCGGAACACAAACCC---AACCCCGCC 50896
 Qy 1831 ThrGlnGlnSerSer-----GlySer-SerGlySerS 1841
 Db 50895 AGAGATCAGGATCCCGACACCCAGAGCGTCCGACCGCGCAGCGCAGCGCTTCCAA 50836
 Qy 1841 rGlyGlyGlyGlySerSerArgProAla-----SerHisSerHisAl 1857
 Db 50835 CCCAGACAAACAGCGCTTACCGTCCGCTCCAGCAGCAGCGCGCGCTCTCTCCAAACGCGGC 50776
 Qy 1857 aHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSer-- 1876
 Db 50775 CCGGGAAGAACCAACGAAACGCGGCATCAAGGGATGAATATCGGGAGATTATCCAA 50716
 Qy 1877 -----ValLeuHisAsnThr-- 1881
 Db 50715 ATAGGACATCAATCGGCTCGCTGAGCGCTTCCGGTGTCTTCCGCGACAGAT 50656
 Qy 1882 -----GlyMetIlyGlyIleThrAlaValGluProSerIlySerProThrValLeu----- 1898
 Db 50655 CAACGGGACCGCTGGGATGTCACCGAGTAGATTCCGATTCGGGACCGTGTCTTCCGCGCAAT 50596

Qy 1899 -----ArgSerThrSerThrSerSerProValArgPr 1909
 Db 50595 CTCGCTCGTGTATTTCGAGGCTTGTTCGAGGATGACATGTGCTGGTGGCTGATGCC 50536
 Qy 1909 o-AlaAlaThrPheProPro-----AlaThrHisCys----- 1919
 Db 50535 GAACGACGACACTCCGCGCCGACGAGGATGGCGCTGTCGGGCCAGGGCGTGTCTCCGT 50476
 Qy 1920 -----ProLeuGlyGlyThrLeuaspGlyValTyrP 1930
 Db 50475 AAGGAGCTGGACCTCCCGCGCGCACCAATCCACTT-----C 50440
 Qy 1930 roThrLeuMetGluProValLeuLeuProLeuGluAlaProArgValAlaAArgProGluA 1950
 Db 50439 CGCGTAGTGTTCATCATCATGCAACGTCGGCGGAGCTGCTGTCGCCGATCGCCATCAC 50380
 Qy 1950 rgProArgAlaAspThrGlyHisAlaPheLeuAlaIlySerProAlaArgSerGlyLeuG 1970
 Db 50379 CATCTTGATCACACGACCAACACCGCTG-----CCGCTGCTGATGACCAATATT 50329
 Qy 1970 luProAlaSerSerProSerLysGlySerGluProArgProLeu----- 1984
 Db 50328 CGACTTCACCGACCCCAACCAACGCGCGATCCCTATCCGCTCTGCTCCCTAAGTGGC 50269
 Qy 1985 -----ValPro--ProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnL 2002
 Db 50268 AATCAGTGTCTGCGCTCAATCGATCACCAAGCTGTGTCAGTCCCATGCGCTCCAC 50209
 Qy 2002 euAlaProHisHisAlaSerProAspProAlaPro-----ProAlas 2017
 Db 50208 AGCGTCCATCAGACACGACGACGCGCCGCACTCGCCAAACGCTGCGTGATCACCGCTG 50149
 Qy 2017 erAlaSerAsp-----ProHisArgGluIlyThrGlnSerL 2029
 Db 50148 CTGCGACGACCATTCGCGCGCGTCAACCATTCGACGCGCACCATCTGTATTCACCGCGT 50089
 Qy 2029 yaProPheSerIleGlnGluLeuLeuArgSerLeuGlyTyr----- 2043
 Db 50088 ACCCGCAGCAGCGCCAGAACCGCGTCCGCTGCGCGCGCATCCGACAGCGCTCCAA 50029
 Qy 2044 -----HisGlySerSerTyrSerProGluGlyValGluProValSerProValSer- 2061
 Db 50028 CAACACGAGCGCGCTTCCGCCCGAGCGGTGCCATCCGACGACGCGCAACACACTT 49969
 Qy 2061 ----- 2061
 Db 49968 ACACCGCGCTCGCGCCGCAAAACCCCTCTGGCGCGGAACCTCCACGAACATCCCGCGCT 49909
 Qy 2062 --ProSerLeuThrHisAspLysGlyLeuProIlyHisLeuGluGluLeuAspLysSerH 2081
 Db 49908 GGCCATCACCGTCCAC-----CACCGCAAGAGCCAGCATCACACTC 49867
 Qy 2081 isLeuGluGlyGluLeuArgProLysGln----- 2090
 Db 49866 ACCAGACCGAGTGTCTTGA-CCGCGCAGGTGATGACACCAACGAAATACGCCGCCGGAAGA 49808
 Qy 2091 -----ProGlyProValIlyLeuGlyGlyGluAlaAlaHis-----L 2103
 Db 49807 TATCCACCGTACCGCAGGACCTCTGAAACCAACGAAATACGCCGCCGCGCAAGAAGA 49748
 Qy 2103 euProHisLeuArgProProGlu-----SerGlnProSerSerSerProLeuG 2121
 Db 49747 CGCGCGCGCGCTGCTTACCTAGATAACCTTCGAAACCTCCGCGCGCTGCTGACGA 49688
 Qy 2121 lnThrAlaProGlyValIlyGlyHisGlnArgValThrLeuAlaGlnHisIle----- 2139
 Db 49687 ACCGCGCGCGTA-GTCGTGTGTGATCAGCCC-----CGGAACACGCGCGTC 49641
 Qy 2140 -----SerGluV 2142
 Db 49640 CGACTCCTCGCAACGAAGCGGATCAATCCCGCGCGCTCCAAAGGTCTCCCAAGCATC 49581

```
Qy 2142 alleleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSer-AlaProLeuProAla 2161
Db 49580 TCCAGCAACACCGCTGCTCGGATCCATCCCAACGCCCTCACGCGGGAAATCCCGAAG 49521
Qy 2162 ProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSer 2181
Db 49520 AAACCGGCATCGAACTCCGCGCATCT- - - - -CGCAGGAAGCGCGCCTCA 49476
Qy 2182 AspleuTyrLeuProProAspHisGlyAla- - - - -ProAlaArgGlySer 2197
Db 49475 CACACATACGAGTCCCGAGACGATCGGATCCGGATCGAACAACCCCGCTGATCCAA 49416
Qy 2198 ProHisSerGluGly- - - - - 2202
Db 49415 CCACGATCGTCCGGGAACCTCCGCACCGCATCCAAAGCCCGCGCACCAACGCAAGC 49356
Qy 2203 - - - - -GlyValArgSerProGluProAsnLysThr- - - - - 2212
Db 49355 TCCTCCGGCGAGGACACCCACCGCGAAACGACACCCCATCCCGATCCGATCAACCGGG 49296
Qy 2213 - - - - -SerValLeuGlyGlyGlyGluAspGlyIleGluProValSer 2226
Db 49295 TCCTCACCACCTGCGCGCTCCACAAACCGCGGTGCCAA- - - - -TTGCGTGCCTCCACG 49242
Qy 2227 ProProGluGlyMetThrGluPro- - - - -GlyHisSerArg 2238
Db 49241 CCACCAATCATGTTGTCCGCGCAAGAACCGAAGCCCGCGGTGTCGATAGTCGAAC 49182
Qy 2239 SerAlaValTyrProLeuLeuTyrArgAspGlyGluGln- - - - -ThrGluProSer 2255
Db 49181 ACCGCGAGTA- - - - -GCCGGAACCGAAGCGCCAGCACCCCGCGCAAGC 49140
Qy 2256 ArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeu 2275
Db 49139 CGTTCCGCACTCCACACCGCAACGATCAAAACCCAAATCCCGGAAAGCAGCTCG 49080
Qy 2276 ThrGluSerAsnSerAlaMetValLysSerLysGlnGluLeuAsnLysLysLeuAsn 2295
Db 49079 ACCCAACCTCGACCGCA- - - - -CCCGAATACCCACGACACGCGGAA 49038
Qy 2296 ThrHisAsnArgAsnGluProGluTyrAsnLysSer- - - - - 2307
Db 49037 ACATGCTCCGCGCAACCAACCAACGACGAGGTTTCTGTATCCAGTCCGGAAGCCTCCGC 48978
Qy 2308 - - - - -GlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGly 2323
Db 48977 AGCGGACCGACGACGCGTGACATCTCCGACTCGTTGACCGGCTCCCGCGCACGGC 48918
Qy 2324 LeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeu- - - 2342
Db 48917 GCGGGCAC- - - - -AGTCTGTCGAACCAACGACGACGACACCG 48882
Qy 2343 - - - - -GluAlaIleAlaArgLysAlaLeuMetGlyLysTyrAspGlnTyrGluGluSer 2360
Db 48881 CGTCCGCGCATCGCGCGCAGCGGCTTCAGTTCGATGTGGACCGGACGAGACGAT 48822
Qy 2361 ProProLeuSerAlaAsnAlaPhe- - - - -AsnPro- - - - - 2370
Db 48821 TCGCCCGGACATGAACGCATCTCGAACGCGGCAACCCGCTCCTCCGCGGATCCGC 48762
Qy 2371 - - - - -LeuAsnAlaSerAlaSerLeuProAlaMetProIle 2383
Db 48761 GGATTGTGCGGAAATGCGGTTCGCTCCGCGCATCGAGCGCTCCGCGTATCCCGCTG 48702
Qy 2384 ThrAlaAla- - - - -AspGlyArgSerAspHisThrLeuThrSerProGly-GlyGlyGly 2402
Db 48701 GTCGGCTCCCAAGCCCAAGCAGCGACGACGACGCGAGCGAGTCCCGCGTCCCGGTCG 48642
Qy 2402 sAlaLysValSerGlyArgProSerSerArgGlyAlaLysSerProAlaProGlyLeuAl 2422
Db 48641 TGAG- - - - -CCAGGCGCGTCCAGAACGCGT- - - - -TCGCCGCGCGGTAGTTTCCT 48597
Qy 2422 aSerGlyAspArgProPro- - - - -SerValSerSe 2432
```

```
Db 48596 TGACCGCGACGCCCCACCCGAGAACGAGGAGAGACGACGAGAGATCC 48537
Qy 2432 rValHisSerGlu- - - - -GlyAspCysAsnArg- - - - -ArgThrProLe 2445
Db 48536 AGTTCACGGGTACGTCGTCGAGTGCCACGACCGCGGACCTTCGGCGCCGACACCTTT 48477
Qy 2445 uThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAs 2465
Db 48476 GCCACGTGCTCTGGTGTCAACGACTCGACGACCCCGTCGCGCAGTGCCCGCGCATGC 48417
Qy 2465 nProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuPr 2485
Db 48416 ACCACTC- - - - -CGCGACGCGGGAACCTTCGCGAGATCGTCTGCAGCAGCGGCTCC 48366
Qy 2485 oAlaGlySerGlyProLeuAlaGlyProHisHisAlaTrpAspGluGluProLysProLe 2505
Db 48365 AGCTGGCGCGGTCCC- - - - -CGACATCAGC- - - - -TCGCCACCTCGACCACA 48321
Qy 2505 uLeu 2506
Db 48320 GCTC 48317

RESULT 86
BX248342
LOCUS
DEFINITION
BX248342 BX248333
VERSION
BX248342.1 GI:31619031
KEYWORDS
SOURCE
Mycobacterium bovis subsp. bovis AF2122/97
ORGANISM
Mycobacterium bovis subsp. bovis AF2122/97
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE
1
Garnier,T., Eiglmeier,K., Camus,J.-C., Medina,N., Mansoor,H.,
Pryor,M., Duthoy,S., Grondin,S., Lacroix,C., Monsemp,C., Simon,S.,
Harris,B., Atkin,R., Doggett,J., Mayes,R., Keating,I.,
Wheeler,P.R., Parkhill,J., Barrell,B.G., Cole,S.T., Gordon,S.V. and
Hewinson,G.
The complete genome sequence of Mycobacterium bovis
Online Publication
PNAS 10.1073/pnas.1130426100 ( Microbiology )
2 (bases 1 to 306550)
Garnier,T.
Direct Submission
Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire
Bacterienne Institut Pasteur 28,rue du Dr Roux 75724 PARIS cedex
15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary
Laboratories Agency Weybridge, Woodham Lane, New Haw,Addlestone,
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
Hinxton,Cambridge CB10 1SA, UK. PT4 Annotation, Genopole, Institut
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28
rue du Docteur Roux, 75724 Paris Cedex 15, France
FEATURES
Location/Qualifiers
source
1. .306550
/organism="Mycobacterium bovis subsp. bovis AF2122/97"
/mol_type="genomic DNA"
/strain="AF2122/97"
/db_xref="taxon:233413"
201. .1286
/genes="adhE2"
/locus_tag="Mb2283"
201. .1286
/genes="adhE2"
/locus_tag="Mb2283"
/EC_number="1.2.1.1"
/notes="Mb2283, adhE2, len: 361 aa. Equivalent to Rv2259,
len: 361 aa, from Mycobacterium tuberculosis strain H37Rv,"
```

(100.0% identity in 361 aa overlap). Probable adhE2.
zinc-containing alcohol dehydrogenase, similar to several, especially mycothiol-dependent formaldehyde dehydrogenase from *Amycolatopsis methanolica* P80094 (360 aa). Contains P500059 Zinc-containing alcohol dehydrogenases signature. FASTA scores: >sp|P80094|FADH AMYME
NAD/MYCOTHIOL-DEPENDENT FORMALDEHYDE DEHYDROGENASE (MD-PALDH) length = 360, Expect = e-156, Identities = 268/358 (74%). Also similar to Rv0162c, (MTC128.02c, 35.0% identity in 371 aa overlap)."
/codon_start=1
/transl_table=11
/product="Probable zinc-dependent alcohol dehydrogenase AdhE2"
/protein_id="CAD97144.1"
/db_xref="GI:31619032"
/translation="MSQTVRGVIAKQGFVVELNIVVPPDGPGEAVVDVTCGVCHTDLTVREGGINDEYFLLGHEAAGIIIEAVGPGVTAPEFGFVILNWRAGCQCRACKRG RPYCDPTNAEOKMTLTGTGELTVAIGAPADKTLVHSGQCTKVDPAADPAVAGLL CGCVMAIGNAINTGVTTRDDTVAVTGGCGVDAAGALGAKBIIAVDTDGDKLD WARTGATHTVNAREVDVQVIGLIDGFGADVVIDAVGRPTYQAQAFYARDLAGTVV LVGPDPDMLDPLDFFSHGALKSKSYGDCLPSPDFTLIDLYLQGRLLPQRFVVS ERIGLEDEVEAFHMGKGLRSVVM"
/locus_tag="Mb2284"
/locus_tag="Mb2284"
/note="Mb2284", len: 211 aa. Equivalent to Rv2260, len: 211 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 211 aa overlap). Conserved hypothetical protein, similar to hypothetical proteins Rv0634c, Rv1637c, Rv3677c, Rv2581c from Mycobacterium tuberculosis and to various hydrolases. FASTA scores: sptr|O06154|O06154 HYPOTHETICAL 21.3 KD PROTEIN (200 aa) opt: 355, E(): 4e-15; (37.4% identity in 198 aa overlap)."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD97145.1"
/db_xref="GI:31619033"
/translation="MAIERVITHGTGTFELDGSGSEVDNNIWLVDGDSVVVFDAAHHA APIIDVAGGRKWAVITGHNDHVTVAPELGTALDAPVLMHPGDVLMWRMTHPKDSF RAYSDGDAVGVGTGELRALHPGHS PGSCVWYAPELGPGTGTGDTTFAGGPGATG RVSVDFTILRSGLGALPGDTVHTGHGDSITIGDEIVHVEEWVARGH"
complement(1998)..3506)
/locus_tag="Mb2285c"
/locus_tag="Mb2285c"
/note="Mb2285c", len: 502 aa. Equivalent to Rv2262c and Rv2261c, len: 360 aa and 140 aa, from Mycobacterium tuberculosis strain H37Rv, (94.7% identity in 357 aa overlap and 100.0% identity in 140 aa overlap). Conserved hypothetical protein, with function unknown but some similarity to N-terminal 70% of P23930|P77703|LNT ECOLI|CUTE|B0657 APOLIPOPROTEIN N-ACYLTRANSFERASE (EC 2.3.1.-) from *Escherichia coli* strain K12 (512 aa). FASTA scores: opt: 239, E(): 1.6e-07, (30.4% identity in 359 aa overlap). Note that neighboring ORF shows similarity to N-terminal part of PCC6803 apolipoprotein N-acyltransferase from *Synechocystis* sp., suggesting possibility of frameshift. Sequence of clones from two sources has been checked but no error found. Appear to be two extra bases at position 1876970 compared to CDC1551 strain. Conserved hypothetical protein, with function unknown but some similarity to C-terminal end of PCC6803 apolipoprotein N-acyltransferase from *Synechocystis* sp. Note that next ORF shows similarity to N-terminal part of P74055 APOLIPOPROTEIN N-ACYLTRANSFERASE from *Escherichia coli* (519 aa). FASTA scores: opt: 142, E(): 0.007, (29.9% identity in 117 aa overlap), suggesting possible frameshift. Sequence of clones from two sources has been checked but no error found.

REMARK-M.bovis-M.tuberculosis: In Mycobacterium tuberculosis strain H37Rv, Rv2262c and Rv2261c exist as 2 genes. In Mycobacterium bovis, a 2 bp deletion (ct-*) results in a single product which is more similar to Rv2262c."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD97146.1"
/db_xref="GI:31619034"
/translation="MALRAGARRQPVIGCAALVFGGLPALAFAPAPSWWLAWFGLVPLLLVVRAPATSWEGALRAWTGNGGFVLAQTQWLVTISAGPMLVLLAAGLGVWLPAGWL AHRLLSVPTTTCVGAALVVVPSVAARAEVSWQSLGPGWALLGASQWSPVTLASA SLGVWLTSTFLVATNATIASVLCVCRATGRLVALGCVIGCAGLSPALYLSVPGVG PTVRVLVQAGDIADAAARLAAGBEFTAAVADQRPDLVWGESVVGQDLTRHPDVILR LAELSORVKAALKRORQAGPVPVLAVALNSLHIAPLISVEMTSDLTRHAARLGAALLVY FGWITRYSKAALKRORQAGPVPVLAVALNSLHIAPLISVEMTSDLTRHAARLGAALLVY QSSTSTFGSNAQPOLAAQAPVAVRAGIPAVHASLSDSSAFDTRGRRLANCSAEFN GAIVNVFLASNVTLVYLRGLDWVPVTFVVMVAGFAVFLRRSLARVSDCAK"
3595..4548
/locus_tag="Mb2286"
/locus_tag="Mb2286"
/EC_number="1.-.-.-", len: 317 aa. Equivalent to Rv2263, len: 317 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 317 aa overlap). Possible oxidoreductase (EC 1.-.-.-), similar to several oxidoreductases. Similarity suggests alternative GTG start at 10154 but then no zbs. FASTA scores: sptr|O544 05|Q54405 PROBABLY AN NADP-DEPENDENT OXIDOREDUCTASE (297 aa) opt: 487, E(): 1.1e-23; (36.1% identity in 299 aa overlap). Also similar to M. tuberculosis Rv0068, and Rv0439c."
/codon_start=1
/transl_table=11
/product="Possible oxidoreductase"
/protein_id="CAD97147.1"
/db_xref="GI:31619035"
/translation="NAKDLAVTVPLSGKLAIITGANSGLGFLARRLSAAGADVIMA IENRAGAAVEEITRTAVPDAKLTIKALDLSLASVAALGQMLADGRPIDLLINNAG VMTPEVTTADGFELOQSGNHLGHATLHLLPRLAAQRAVVSLSLAARRGRHIF PDLQERSYAPMTAYGSKLAVLMEFALDRSRSAAGWGLISNAHPLGLTNTLQIA GSHGRDIPALMERLYKTSWRPAPFPEIEIRGILPALLYAATPOADGAGFYGRGRY EVAGGVEAKVPAARNDANCKRLMEVSEQLTGVSYPKSR"
complement(4526)..6304)
/locus_tag="Mb2287c"
/locus_tag="Mb2287c"
/note="Mb2287c", len: 592 aa. Equivalent to Rv2264c, len: 592 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 592 aa overlap). Conserved hypothetical Pro-rich protein, similar to hypothetical proteins Rv0312 (MTCY63.17, 620 aa and Rv0350) that has highly Pro-, Thr-rich C-terminus. Contains P500343 Gram-positive cocci surface proteins 'anchoring' hexapeptide. FASTA scores: Z96800|MTCY63.17 Mycobacterium tuberculosis cosmid (620 aa) opt: 1075, E(): 8.8e-24; (38.9% identity in 627 aa overlap)."
/codon_start=1
/transl_table=11
/product="conserved hypothetical proline rich protein"
/protein_id="CAD97148.1"
/db_xref="GI:31619036"
/translation="MATGAPALGLSIGVTNLAANAADHSITRKPVLTLYRPPPEVG VPSENRLDEPGLVITDFVDRVGDVSVIADGVSVRSEALVADALLAYATGGRAD LRGSVTYPAHWGPAVALDALSARASWESHGTSSTAQPLSLPDAALYAIARAD PGPARGIYAVCDFGSGGTILVDADRYVPVAAVTRHQAFSGDLIDOSLLSYVNSE LPQTGAPDAGTSAIGSLTKRLECKRERLSVGGANIPAVTITLGGSDILRLNRELDT IRDLSVSGRALEQTLRSGIRGIAELVAIVSUGGANIPAVTITLGGSDILRLNRELDT PQLTAARFGALWAARPSTATVATVATATAPADAPASVLOPALAWSEADESH IGPAQVTAARPSLSFDHAAPEPEKSPPIPWYRLPAVITGTTVALLVGAIVAIG LSTGDTQPTAGTPQRPQVTTTAAAPPSPAPASDGTTEPAPPVQAPATGGPAPLQOP

LPPPTTTTQPAVTTDVITPAPTTPASAPPATTQPPATTQPPATTSPSPPTPIPP
 IPIPOLPPGIPQVPGIGQFSAISGS"
 6654..7883
 /locus_tag="Mb2288"
 6654..7883
 /locus_tag="Mb2288"

gene

CDS

Alignment Scores:

Pred. No.: 0.00024 Length: 306550
 Score: 666.50 Matches: 629
 Percent Similarity: 32.73% Conservative: 291
 Best Local Similarity: 22.38% Mismatches: 1094
 Query Match: 5.04% Indels: 814
 DB: 1 Gaps: 127

US-09-522-753-5 (1-2517) x BX248342 (1-306550)

Qy 20 ProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeu 39
 Db 254553 CCCCCGACGGGTTCAGCGC-CCGCGCAGCAGGCGCGG-----ATC 254593
 Qy 40 LeuGluTyrGlnHisHisSerArg-----AspTyr 49
 Db 254594 CTCGATGTCGGCGCTATAGCCAGCGCTTCTGGGCATCATCTAGGCGGTGCGTGATGCC 254653
 Qy 50 AlaSerHisLeuSerProGlySerIle-----IleGlnProGlnArg 64
 Db 254654 GGCCACGGCATCTATCAGGCGCGGTGCGGCAGTGCAGCGCATACACACAGGTT-CGG 254712
 Qy 65 ArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeu 84
 Db 254713 CCCCCCGGATGTGTCATCGTGTGTCGGCCAGGCGCGGCGGCAACAGCGCCCTCTT 254772
 Qy 85 -----ArgProGluSerHisSerTyrLeuProGluLeuGlyLysSer 98
 Db 254773 GCAGGCGCGCGCGGACGCCAACAGCGCTGCAGACATGATGCCAACAGCAACCGGTGGC 254832
 Qy 99 GluMetGluPheIleGluSer-----LysArgPro-----ArgLeuGluLeu 112
 Db 254833 CGGTATCGCCATTCTCGCGGTGAACGGAAGCGCGCGGTAGTGTGTCGATCTCGA-- 254889
 Qy 113 LeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySer 132
 Db 254890 -----GCTCGGCTTGTGACCGCGCTTTCATGCGCGCGCCACCGAGGGG 254934
 Qy 133 GluAspLeuThrLysAspArgSerLeuThrGly-----Lys 144
 Db 254935 CGTCAAGCGCGCGGCAACGCGCGGTAGTGTGTCGCGGTGTCGCGCACGACGCGATCGG 254994
 Qy 145 LeuGluProValSerProProSerProHisThrAsp---ProGluLeuGluLeuVal 163
 Db 254995 CATCGCGGATTACAGCAGCTTCTCCAAACGATACCTGGCGCACCGTCTTAGCAGGCGGT 255054
 Qy 164 ProPro-----ArgLeuSerLysGluGluLeuIleGlnAsnMetAspArg 178
 Db 255055 AGCGGTTTCGCGCGTGTGTCGTCGTCACACACAGAGAGTTGTCGCCAGCAGGCTGA 255114
 Qy 179 ValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGln 198
 Db 255115 GCAGATCGACACCTCGTAGCGTGAACAGTCACCGCGCGGCGGTGCGGCTTGGCGACCGT 255174
 Qy 199 GlnLeuGluGluAlaAlaLysProGluProGluLysProValSerProProPro 218
 Db 255175 CGAGATCAACCCGCTCGGGAAACCGCAGTCGCCGAA-----ACAGCACCTGTCTCCG 255228
 Qy 219 IleGluSerLysHisArgSerLeuValGlnIleIle-TyrAspGluAsnArgLysLysAl 238
 Db 255229 GTC-----CGGTGAGCAGCGCATGTGACCATGTCAGTCGAGCAAGCCGCGATCGTCT 255276
 Qy 238 aGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAs 258
 Db 255277 GCTGGCGCGCGACCGCAATACGCGATCCACCGGTGACAGCGCGGAA---CCGGTCATGCA 255333

Qy 258 nGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLy 278
 Db 255334 AGCTGTGACGATTTTCGTCAG----- 255355
 Qy 278 sLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPhe 298
 Db 255356 -----CGCCAGGCGACGCCCGCGCAGC----- 255379
 Qy 298 eCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGlu----- 314
 Db 255380 -TGC--AAGTTGATCGCAGCGGATGCGGTGCGTTCGTCGATCTCGGTCCACCA 255435
 Qy 315 -----ArgIleGluAsnAsnProArg----- 321
 Db 255436 GGGCGAGGTTGTGCGCAGTGTATCTCGAGTTTCGGCGCGCGCTCACGACGCGGTGCGGTGA 255495
 Qy 322 -----ArgArgAlaLysGluSerLysValArgGluTyrTy 333
 Db 255496 ACAACTCGATCGCTCGCCGTGCCAGCGGGGAAACCCGCCAAATCTGCTCACCGGCCA 255555
 Qy 333 rGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerAr 353
 Db 255556 CCGGATCGGTTCCCGCTGCTGCGCAATACCTCAGCGGTGGGACGCCCGGAGCAACG 255615
 Qy 353 gValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHis---GluVa 372
 Db 255616 CGACGATCAGAGCGCGCACCGCTCGAGCAAGTCTCGCATGTTCTCCAGCACTACCAACA 255675
 Qy 372 lSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLe 392
 Db 255676 TGGCGCGGTGCGCGAT-----ACGCGCAATATGGTGTCCACGTCAGCGCGCGGTGATCG 255729
 Qy 392 uAlaValIlePro-----ProMetLeuTyrAspAlaAspGlnGlnArgIle----- 407
 Db 255730 GATCCGCAACCCCAAAACCGCGCGCGCTCGGCATCGGCACCGCGCGGTGATCG 255789
 Qy 408 -----LysPheIl 410
 Db 255790 GCGCAGGTTGACATACCAAAACCCGTCGCGATAACCGTCGCGCAACCGCGCTCGCACT 255849
 Qy 410 eAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAs 430
 Db 255850 GTGTGCGCAGCGGTGCTTTCGACACCCCGCGCACCGGTAAAGGTCACCCACCGTTGA 255909
 Qy 430 nMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAs 450
 Db 255910 CGTC-----CAGCAGCCACGAGTTCGCGCATCTGTCACGCGCGCCACCA 255957
 Qy 450 nPheGlyLeuIleAlaSerPheLeuGluArgLysThrVal----- 463
 Db 255958 G-----CCGAGTCAGCTGGCGCGGAGACAGATGCGCACCAACGACTTTCGGG 256005
 Qy 464 -----AlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyrLy 480
 Db 256006 TCCGACGCGCGGGAACCGCTTGTGCAGATCAGGGTCACACAGCTGACCAACCGCTTCCG 256065
 Qy 480 sSer-----LeuValArgArgSerTyrArgArgArgGlyLysSerGl 494
 Db 256066 GTCGGGCAGGTGTCGACGCGGTAGTACCGAGTTCGTCAGCAGCGCTCTTGGGCA 256125
 Qy 494 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 514
 Db 256126 GCAGGTGAGCAACACAGATCGTGTGTAGTTCGCGACAACACGCTGCGCGCGCGTGGGCA 256185
 Qy 514 gSerSerGln----- 517
 Db 256186 GCTCGCGCACCGCGCGGTGCGGTGCGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 256245
 Qy 518 -GluGluLysAspGluLysGluLysGluLysGluLysGluLysGluLysGluLysPro-- 536
 Db 256246 ACACACCTCCCGGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 256305
 Qy 537 -GluValGluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGl 556

```
Db 256306 GCAAGCCAGGGGCGACCGCACGGCGTC-----GGATGCGCGGGCGCA 256347
QY 556 uAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThr-----Al 572
Db 256348 ACCGACCAAGAGTGTGCGCTTCGCTTCGACCGGGCAACCCCGCGGTGCTCGC 256407
QY 572 aAsnSerGlnGlyArgLysGlyArgLysThrArgSerMetAlaAsnGluAlaAsnSe 592
Db 256408 GAACCAATTCGGTCAGCGTTC-GGTCAGTTGCGATCGCGCTGTCTCAAGCTGAGAC 256466
QY 592 r-----GluGluAlaIleThrProGlnGlnSerAl 602
Db 256467 CCGCGAGGTTGGTTCGGCCCTCGATATCGGCCAGCAGCAAGCTACCGTCCCGTCGCT 256526
QY 602 aGluLeuAlaSerMetGluLeuAsnGlnSerSerArgTrpThrGluGlu----- 618
Db 256527 ACAAGCTCGCTCACCATCTGCGCTCCAGTCCACAGGT-ACCACGTTCGACGCGCGGGTG 256585
QY 619 -----GluMetGlnThrAlaLysLysGlyLeuLeuGluHisGlyAr 632
Db 256586 AATCTTGCTCATGTAACCAAGCATCGAGCCAGCGGTAGCGCATTAATCGGCAC----- 256640
QY 632 gAsnTrpSerAlaIleAlaArgMetValGlySerLysThrVal-----SerGlnCy 649
Db 256641 -----CTGCGCTAGATTCTCTGAAATCTCTTGGCGCGCGGTCCATGTG 256684
QY 649 sLysAsnPheTyPheAsnTrpLysLysArgGlnAsnLeuAspGlu----- 664
Db 256685 TTCTACGCGCTTAGTCGATCGATTCGCGACCGCGGTGCGATCGCGCGGACCTACAGT 256744
QY 665 ----IleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLy 683
Db 256745 GCGCGTCTCGCTCGCAATTGTGAGTCCCCAGTGTGGCAGCATCGCCCGCGAAGAA 256804
QY 683 sLys-----LysAlaProAlaAlaAspGluGlu 694
Db 256805 CCGACACGCGCATCGTGGCGGTGCGTTCGAGTTCGCGGTGCGCGGTGCGCGAGT 256864
QY 694 la-----AlaPheProProValValGluAspGluGluMetGluAlaSerGlyV 710
Db 256865 CACGCGCGCATGGGTTCGAAAGGTGATGGAATGATGATGATGATGATGATGATGATG 256924
QY 710 al-----SerGlyAsnGluGluGluMetValGluGluAlaG 722
Db 256925 CCGCTTCGCGGTAAACCCCAACCATCGGAAACAGAGCTAGCTCGCCCGCGCGCGCA 256984
QY 722 luAlaLeuHisAlaSerGly---AsnGluValProArg-----GlyG 735
Db 256985 GAGCGAGCATCGCGGTGGCGCCCAACGACGAGGAGCATCGGATTCTCTGACCATGGTG 257044
QY 735 luCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProH 755
Db 257045 GGTGTCTTGGTACGTACGTTA-----ACGGAGATGAGGGGCGCGCTTCGCGC- 257094
QY 755 isThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProAlaThrLeuGlyA 775
Db 257095 -----CTTCAACCGCG----- 257106
QY 775 laAspGlyProProGlyProProThrProProArgArgThrSerArgAlaProIleG 795
Db 257107 --GAACCGCGGAGCGGTTCGCTTCGCGTTCGCGAGCCACCGACTCGCGCGACG 257164
QY 795 luProThrProAlaSerGluAlaThrGlyAlaProThrPro-----ProAlaProp 813
Db 257165 AGCGAGCGCGCGGAGGTCAAGCCACCGGACCGCGGTGCGCGCTCCTCCGCGCGCGC 257224
QY 813 roSerProSerAlaProProValValProLysGluGluLysGluGluGluThrAlaA 833
Db 257225 CGTCCCCCGCTACCGCGCGCGGTGCG-AGGCGGAGGGGCGCGGAGCCCGCGAG 257283
QY 833 laAlaProPro-----ValGluGluGluGluGlnLysProProAlaAlaG 849
```

Qy	1171	nLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnI	1191	Qy	1529	leValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaP	1549
Db	258261	GCCTCACCAATGTCGCGCGCTCACCGCC	258308	Db	259021	-----GAACACGAGCGCGCGAATTTGGCCCGCTGGCGCGCGCGCATCGCCAC	259072
Qy	1191	uAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysG	1211	Qy	1549	roPheAlaGlyHisLeuProArgGlySerProValThrMetArg	1566
Db	258309	GGCCCC-GCCATCACCGCAGAGCTTCGCGCGCTGCGCGCGCCCATCACCGCG	258367	Db	259073	CA-----TTGCCCGCGCGCCACCATCGCCGAGCCCGCGCGCATTTGCCCGCG	259120
Qy	1211	yIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHi	1231	Qy	1566	hrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuT	1586
Db	258368	CTCGCGCTCCCGAATAGCCCGCGCTTCGCGCGCTCACCGCCCTGGCGCGCTCGCGC	258427	Db	259121	CGCTCCGTTCGCC	259135
Qy	1231	sGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleGlyGluAspSe	1251	Qy	1586	hrSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHisHisProH	1606
Db	258428	CGCGCGCGCGCGC-----CTTGCAACCGTTGC-----GC	258457	Db	259136	CGCGCGCGCGCGCGCGCTGGCTTGGCACCGCTTGCACCGCTTACACCGCTTGC	259195
Qy	1251	rProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrG	1271	Qy	1606	isProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgS	1626
Db	258458	CGACGCGCGCTCGCGCGCGCTCAGTGGCGCGCTTGTGCTGCGCTCCACGCGCTT	258512	Db	259196	TGCCG--GGCCCGTGG-----TCGCTGGCGTCCACACCGCTG--GGCG	259234
Qy	1271	uGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLy	1291	Qy	1626	erHisIleProLeuAlaPheAspProThrSerIlePro-ArgGlyIleProLeuAspAla	1645
Db	258513	GGCGCGGAGGTGCGTTCCTGCTCACCCTCCAGACCGCGCGACCGCTTGGCC--	258569	Db	259235	CGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	259291
Qy	1291	sGluAspGlyArgSerSerGlyProProHisGluThrAlaAlaProLysArgThrTy	1311	Qy	1646	AlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyr	1665
Db	258570	-----GGGTCACCGCGCTTACCGCGCGCTCCGC	258598	Db	259292	CACCGCGCTCACCGCGCTCCCGCCATCA-CTGCCATGACCGCGCTCACCGCGCTTGC	259350
Qy	1311	rAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMe	1331	Qy	1666	Pro-----ProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsn	1681
Db	258599	CGCGCGCGCGCGGATACCAATCCGCGCTT-----GCCGCGCGCGCGCCACCGCGC	258655	Db	259351	CCGGCTCCCGCGAGCCCGCC-GTCAACCGCCACACCGCGCGCGCGCTGCTGCTG	259400
Qy	1331	tGlyArgAlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHisIleI	1351	Qy	1682	ArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThr	1701
Db	258656	CGCGCGCGCGCA-GTCCGCGCTCGCCACCGACCCACCGTCCGCTGCTGCTGCTG	258709	Db	259400	-----	259400
Qy	1351	eArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnAspTyrLe	1371	Qy	1702	AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAla	1721
Db	258710	-----CGTCCCGAAGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	258744	Db	259401	-----GCCACCTGACCGCGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	259442
Qy	1371	uArgArgGluAlaLysLeuLysArgGluGlyThrProProProProProSerAr	1391	Qy	1722	LeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuPro	1741
Db	258745	TCCCGCGCATGTCCACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	258803	Db	259443	-----ACCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	259454
Qy	1391	gAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHi	1411	Qy	1742	ValLeuVal-ProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLe	1761
Db	258804	ACGGTCCCGCATCTGTCGATA--CGTTGGCGCGCGCGCGCGCGCGCGCGCGCG	258843	Db	259455	-----GCCGTACCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	259491
Qy	1411	sGlyGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgG	1431	Qy	1761	uProThrAlaProGlnProPheSerSerArgHisSerSerProLeuSerProGlyG	1781
Db	258844	-----GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	258854	Db	259492	-CCTTTCCCGCGCGCGCGCTCCAAACCGCTCAGCGCTGTGGCGCGCGGAGCGCGCG	259545
Qy	1431	uGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGlyGly---	1450	Qy	1781	yProThrHisLeuThrLysProThrThrThrThrThrThrThrThrThrThrThrThr	1801
Db	258855	ACCGCTAGCGCCACACCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	258910	Db	259546	-----TTGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	259574
Qy	1450	xIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSer---ThrThrGlySerLy	1469	Qy	1801	pArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThr	1821
Db	258911	CGTACCCGACACCGACCC-----ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	258956	Db	259575	-----ACCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	259621
Qy	1469	sLysHisAspValArgSerLeuIleGly-SerProGlyArgThrPheProProValHisP	1489	Qy	1821	lGluHisAlaProIleTyrArgProGlyThrGluGlnSerSerGlySerSerGlySerSe	1841
Db	258957	-----GCCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	258994	Db	259622	CGGACACCGACCC-----ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	259654
Qy	1489	roLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuL	1509	Qy	1841	rgLysGlyGlyGlyGlySerSerSerArgProAlaSerHis	1854
Db	258995	CC-----	258996	Db			

Db	259768	CCGCCAACCGCGCTTGACCGCGGTGGCTCGCTTCCTTGGCCCGGTGGCCCGAGAGG	259827	260688	CGCCAGACCATTTGGCCCGCGCGGAGACCGCGCGCCCGCCGAGCGCCC	260746
Qy	1885	YlleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSerSe	1905	2188	oAspHisGlyAlaProAlaArg--GlySerProHisSerGluGlyGlyLysArgSerPr	2207
Db	259828	GTGCGG--TCGGCG--CCGTCCTCACCTAGTTC--GC	259859	260747	GGCTCGCGCAGCCCGCCACCGACCGCGGTGCGCCACCGTGACCGCGCGCCCGCTCACC	260806
Qy	1905	rProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLe	1925	2207	oGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIle--GluProValSe	2226
Db	259860	GCCATTGCGACCGCGCGCGCGGTACACCGCTGCGCGCGCGCCCGCGCTCACCAGCAA	259919	260807	GGCGCGCAGCCAACTGCGCACCGCTTGGCGCGGCACCGCGCTCACGAGGCGTCCACCCAG	260866
Qy	1925	uAspGlyValTyProThrLeuMetGluProValLeuLeuPro-----LysGluAl	1942	2226	rProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyProLeuLeuTy	2246
Db	259920	CCCGCGTTACACCGCGCCCGCTTGCCTGCGCGCTTGCGCCACGCGTGGCGTCCGC	259979	260867	CGCCCGACCGCGCGCGCGCGCTCTCCGCGCGCG--CCGACATACCCC-----	260915
Qy	1942	aProArgValAlaArgProGlu-----ArgProArgAlaAspTh	1955	2246	rArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSe	2266
Db	259980	GCCGTTACCGCCCTTGCGCGCGTTCGCCACCATGTGCAGCGTAGCGCGTTCGCGCGGT	260039	260916	-----AATAGTCCAGCGCAGCCCGCC	260935
Qy	1955	rGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerPr	1975	2266	rGlnProProAlaPheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLy	2286
Db	260040	CACACCCCTCGGTGCGCGCGCGCGCGCGCGG-----CCACCTTCCACCC	260084	260936	AGCACCCCGCGCGCC--GCCACCGCGCGCTTCCGCGCGCACCGCATTTGCGCGCGCGC	260994
Qy	1975	oSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAl	1995	2286	sLysGlnGluLeuAsnLysLeuAsnThrHisAsnArgAsnGluProGluTyAsnIl	2306
Db	260085	GGCGCTGGGTACACCATCGCGCGCGCGCGCCACCG-----TCACCATCTTGAAGC	260138	260995	CGCGCTCACCGCGCGCGCGA-----GATACCGCGCGCGCCCATTTATTCC--	261040
Qy	1995	aArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaProPr	2015	2306	eSerGlnProGlyThrGluLeuPheAsnMetProAlaIleThrGlyThrGlyLeuMetTh	2326
Db	260139	CATGTGCGCGCGCGCACCGCTGCGCGCGCGGTACCGCGCGCG-----CCACCGCGCGCC	260192	261041	-GGTAGCCCGCGCGCACC-----GCCGCGCACCGCGCGCGCAC	261075
Qy	2015	oAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnI	2035	2326	rTyArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIl	2346
Db	260193	GGCACCAACCGTCCCGCGCA-----ACACCTGTG-----	260220	261076	CGGCACCGCGCGCGCGCACCGCCAAACCGCACCGCGCGCGCGTGTACC-----	261127
Qy	2035	uLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTySerProGluGlyValGluPr	2055	2346	eArgLysAlaLeuMetGlyLysTyArgGlnTrpGluGluSerProProLeuSerAlaAs	2366
Db	260221	-CTGCCCGCGCGCA-----CCCGCAGTGGTCCGTACC	260249	261128	-----GAGAGCCACCTCCCGCGCACCGCT	261153
Qy	2055	oValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGl	2075	2366	nAlaPheAsnProLeu--AsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAl	2385
Db	260250	GCCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	260303	261154	TGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	261213
Qy	2075	uGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProVally	2095	2385	aAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysVa	2405
Db	260304	ACCGTTCGCGCGGTGATAGACGACGACGATGAAACCGTGGAGTCCGTACCA-----	260358	261214	CGGC-----ACCGTCTCGCACCGCGCGCAACACCGCGCGCGCGCGCGCGCGCT	261255
Qy	2095	sLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSe	2115	2405	lSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGly--	2424
Db	260359	-----CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	260408	261256	GACCGCGGTAGCACCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	261315
Qy	2115	r-----SerSerProLeuLeuGlnThrAl	2123	2425	----AspArgProProSerValSerSerValHisSerGluGlyAspCysAsnArgArgTh	2443
Db	260409	GCCCGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	260468	261316	CATGACCAACACCGCGCGGTATACCGCC--CGCGCGCGCGCGCGCGCGCGCGCGCG	261372
Qy	2123	aProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIl	2143	2443	rProLeuThrAsnArgVal--TrpGluAspArgProSerSerAlaGlySerThrProphep	2463
Db	260469	ACCGGGGTACCGGATGAAACCGAGAGATCGCC-----GTTCAT	260507	261373	CGGCACACCGCGCGCGCGCTTCCCGAACAGAGCGCGGTGCCCGCGCGCGCGCGCGCG	261432
Qy	2143	eThr-----GlnAspTyrTh	2148	2463	ro-----TyrAsnProLeuIleM	2469
Db	260508	ATCGGCATAGCGGCACCGCGCGGTCTCGTACCGCCCATGACCGCGCGCGCGCGCGTACC	260567	261433	CACACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	261492
Qy	2148	rArgHisHisProGlnGlnLeuSerAlaProLeuPro-----	2160	2469	etArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerG	2489
Db	260568	GACCAACCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	260627	261493	CGCGGTACTTCCGGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	261552
Qy	2161	-----AlaProLeuTySerPhePheProGlyAlaSerCysProValLeuAspLe	2176	2489	lyProLeuAlaGlyPro	2494
Db	260628	GTGGGCGCATGCGACCTTTCGCGCGGTTCGCGCGGTGACCGCGCGGTTCGTCGCGCG	260687	261553	ACCGCGCGCGCGCGCGCG	261569
Qy	2176	uArgArgPro-----ProSerAspLeuTyLeuProProPr	2188	RESULT 87		
				AF210249		
				77457 bp	DNA	linear
				BCT 29-AUG-2000		

DEFINITION Streptomyces verticillus bleomycin biosynthetic gene cluster, complete cds.

ACCESSION AF210249 AF149091

VERSION AF210249.1 GI:9937210

KEYWORDS

SOURCE Streptomyces verticillus

ORGANISM Streptomyces verticillus

REFERENCE 1 (bases 1 to 74593)

AUTHORS Du, L., and Shen, B.

TITLE Identification and characterization of a type II peptidyl carrier protein from the bleomycin producer Streptomyces verticillus ATCC 15003

JOURNAL Chem. Biol. 6 (8), 507-517 (1999)

MEDLINE 99352421

PUBMED 10421758

REFERENCE 2 (bases 1 to 77457)

AUTHORS Du, L., Chen, M., Sanchez, C. and Shen, B.

TITLE An oxidation domain in the BimII non-ribosomal peptide synthetase probably catalyzing thiazole formation in the biosynthesis of the anti-tumor drug bleomycin in Streptomyces verticillus ATCC15003

JOURNAL FEMS Microbiol. Lett. 189 (2), 171-175 (2000)

MEDLINE 20389599

PUBMED 10930733

REFERENCE 3 (bases 1 to 77457)

AUTHORS Du, L., Sanchez, C., Chen, M., Edwards, D.J. and Shen, B.

TITLE The biosynthetic gene cluster for the antitumor drug bleomycin from Streptomyces verticillus ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase

JOURNAL Chem. Biol. 7 (8), 623-642 (2000)

MEDLINE 20500448

PUBMED 11048953

REFERENCE 4 (bases 1 to 77457)

AUTHORS Du, L., Sanchez, C., Chen, M., Edwards, D.J. and Shen, B.

TITLE Direct Submission

JOURNAL Submitted (01-DEC-1999) Chemistry Department, University of California at Davis, One Shields Avenue, Davis, CA 95616, USA

COMMENT On Aug 29, 2000 this sequence version replaced gi:5326869.

FEATURES

source

1. .77457

/organism="Streptomyces verticillus"

/mol_type="genomic DNA"

/strain="ATCC15003"

/db_xref="ATCC:15003"

/db_xref="taxon:29309"

complement (1199..2185)

/codon_start=1

/transl_table=11

/product="SyrP-like protein"

/protein_id="AAG02342.1"

/db_xref="GI:9937211"

/translation="MTGSVLTLPGLGIIIPRREGLTGTGAEDYDLGADGPDWVRAH
TSPRELRATDGLILHGLPTDGDGVHVVSGVGGDPLPYTERSTRVSVKGIY
GLEYPADQIPHNENSAAHWPSLYFFCHTAPDTGGATPIADGRAVLDLIPAEVR
RSQGVVYTRFRADWGLSWQAFOTEDRGDVHCRAGHQEFQSWDGVLRTRHRP
ATAVDPTGAEVWFNOAHLFHPSSLDPDLRQVLELYGENGLPRDLAFADGTPIPAD
LATVRAAYTAALALPWREGDIMLVNLRNMAHGREPFTGERRVLVAMTSADS"

complement (2182..3918)

/codon_start=1

/transl_table=11

/product="peptide synthetase NRPS12"

/protein_id="AAG02343.1"

/db_xref="GI:9937212"

/translation="MTLTLDADFDAQPTPDHAHVHGTVMVYRELELRAGRMART
LAARGAGPTLVAVRIPGPEPVAALLAVLTGAGVPLADDDPPDCRHLDDCAA
LILAEHPSRDRGLTPDEALAPAPFDAAFPVAGDPAYVITYTSSGSRCKGLVEQGA
LGAVLAQARAYDGLSRTVLHSSLSFDMAVTSLMGPLVSGGAIHVLDLKAIASGTQP
PPAASRPTKPTPSHLPLGLLPDCLPTGQLVIGENALTGSLGAPWRAHPDVTY
VNEIGTEATVGCATVREGDAVDGAPFGRFAGTRLDYLDADGEPVAVGVGVEL
HIAGDQALRGYLRGLTERFFDPFADGSRMYRTGDLVRERPDGLEYLGRADGQ
VKVSGYRIEPTGEAVLRGHAGVRDCAVAVGADARRLVAVYVVPDPSPPGTAAAPAR

CDS

complement (7859..9214)

/codon_start=1

/transl_table=11

/product="putative transporter"

/protein_id="AAG02348.1"

/db_xref="GI:9937217"

/translation="MMKSRLRDRQLGGEDPVVAQBSPDQAGTFCQGGDGLNVFAAL
AALLEVEVPVRLPHKSHGLRHVHSSQAAVALGPMQVATTTGAGDGHQAGRRGQ
VTGVGVREAVAGDDDELGAEDRPHARQLRDLGLMMAERLADLLNPLQTVVQGDL
RGQVSHDLGDLVLPQGRGLSLGSLQRRRGDVGTAHVSQGPGRQFPGPAAAESCRG
LVAROODQALLRAVVEGPFQGEDAGQAVETVDPDVPVGEVGTGGQOGL
GHVDRREPVSVAGGDDVGVAGVGLRLAPVSAHVAADRAAGHVGLGLAVGREGQQQ
QKGRAGDVHCPVDLVVGQVEDLTHGGQDRRLIVADLLRQORHTLVDRHPRVMPVLA
RQVDPGPGSLAHSPLLTPNSIPSTRGTPTCHLKKRKRKTSQSAARAPRTRGRPLVVS
H"

complement (9211..17736)

/codon_start=1

/transl_table=11

/product="peptide synthetase NRPS11-10"

/protein_id="AAG02349.1"

CDS

complement (3915..4733)

/codon_start=1

/evidence="not experimental"

/transl_table=11

/product="hypothetical protein"

/protein_id="AAG02344.1"

/db_xref="GI:9937213"

/translation="MSKNRAHWSRITGDAPGVVLAVDVFGTGRQETAFHLCDLLTD
PVEVHAPVAPDGDWSTATGAGHLRWTEGLDVLGRVPRVALVGCAGGVFASAIRA
DALVERGHRPRVLPFNPSAGVATLTRDFRGLIAGMDLLTDGERAALLAETTAIRLA
HAPDALVPVAERYAALYREGCDLLCERLGVDSFGAELAAVLHLSYLAYLTAALDVPTT
PLMRGAVSLTSTREHQGTFTDVEHGFDAEALLSSPVVVAALTALLREHEASR"

complement (4741..6024)

/codon_start=1

/transl_table=11

/product="putative transporter"

/protein_id="AAG02345.1"

/db_xref="GI:9937214"

/translation="MTARVDATPTYLAVALAVREARAPLLGSLCLARMSFAVLPLALLLS
VRDATSPVAGLTSGALSATLTLPAPARALIDRGSRGLVRLTVPYLLGLAVLIT
LAEAEAPTALLVAAAVAGVAPPGLPTMRWLRIHGRQLHTAIYADLSVTEEVV
FTVPLLAGGLTAAAPLASMITVMVLIAGTACFVLSAATAAPASGEADEDRPHGR
PMALPGMRTIVISFGVGLVGVLPVLPFIADHAGSPGAGGILLMSLGSAGVGGLA
YGRIAWRSTPVRFRVLTGFTLAVPLCLTASVPAGAPALLVGLCLAPLFTTAYLL
VNDLVTASGTAPTEANTWSTANNGGFAGSAAAGVLLDSRGFTVTVTAAFAVAAAAT
VMVLVRRLLTGAGHPPEAATPADRTAPAEAB"

complement (6021..6773)

/codon_start=1

/transl_table=11

/product="thioesterase"

/protein_id="AAG02346.1"

/db_xref="GI:9937215"

/translation="MDFLTRVNPWFSGGCDGRPRVRLCALPYAGGTAAVFKWPAAL
PGVELLTAHLPGRGDFTTEPPATLEAEARLCEALPPSDLPVVLHSGMALLYGE
VAARLAARGAPNLIIAACRPHPVDPDASGPVTEAEALATRAERPWDLTALRDEELM
EALVPALVADITAGDRYHRPRPLDPLKVIYICADDGDTWRTTLGWRACRATDCEV
VVLPGGHYFLETDRAAVLRATVADLAEAEVGA"

complement (6784..7797)

/codon_start=1

/transl_table=11

/product="thioesterase"

/protein_id="AAG02347.1"

/db_xref="GI:9937216"

/translation="MTDLPLRTVALTGESAEVDDLLRLTLADVPDSTVGLLHRLRLA
AQELPURIARLTMRLYDPSRALVVTGVDDEIRIGPTTAPAPADPPERTDLELL
LLHAALLGEAFGATQOGRGLVHDVLPVGEETAQMGSSSETELLWHTEDAFHLRCD
YVGLLCLRNHQRAATTVGWPDLSRLTTEDRAVLEPRYLIRPDTSHTPAQATGTRSA
ERFAAIAEMDDAPERVAVLFGDPEDPYLRIDPAYMSPAGDAAARAYDTVTALIEDE
LRHVLDAGSLLLVDNYQAVHGKRPFAAYVDRGRWLKRVNITDLRRSRSSARRSAT
S"

complement (7859..9214)

/codon_start=1

/transl_table=11

/product="putative transporter"

/protein_id="AAG02348.1"

/db_xref="GI:9937217"

/translation="MMKSRLRDRQLGGEDPVVAQBSPDQAGTFCQGGDGLNVFAAL
AALLEVEVPVRLPHKSHGLRHVHSSQAAVALGPMQVATTTGAGDGHQAGRRGQ
VTGVGVREAVAGDDDELGAEDRPHARQLRDLGLMMAERLADLLNPLQTVVQGDL
RGQVSHDLGDLVLPQGRGLSLGSLQRRRGDVGTAHVSQGPGRQFPGPAAAESCRG
LVAROODQALLRAVVEGPFQGEDAGQAVETVDPDVPVGEVGTGGQOGL
GHVDRREPVSVAGGDDVGVAGVGLRLAPVSAHVAADRAAGHVGLGLAVGREGQQQ
QKGRAGDVHCPVDLVVGQVEDLTHGGQDRRLIVADLLRQORHTLVDRHPRVMPVLA
RQVDPGPGSLAHSPLLTPNSIPSTRGTPTCHLKKRKRKTSQSAARAPRTRGRPLVVS
H"

complement (9211..17736)

/codon_start=1

/transl_table=11

/product="peptide synthetase NRPS11-10"

/protein_id="AAG02349.1"

CDS

CDS

CDS

CDS

CDS

CDS

/db_xref="GI:9917218"

/translation="MASDALSAHLEUSDCORRGGLTIVTNDGSELGAQNVAAVRPERYS
ALAPERTAALHGAATGYDELNRRRAELTRRLADAGAGSTLVAVALPRDPLVATLC
ALLKGAACLPDGLPGIAGRLREIMADASPDLVITRAVAPAFDGGPVLFDAPPT
CVAFLPHSAGTATAYVLYPTTDEKSENSVSDRYMARDLDDPTAGIPARAILR
LVAFLPSGLVLDADDETTPRVTREAPRDMVDVVAQVCAVLGVVGRDRFDL
GKSLAAOVAVLRKLLGVELPLRALPDATVEELARVAEQAGGOGVREAALEP
VGRSEPLPSAQORLWFLDRLMPORAFYTMCDARVGGIDGLRALRLMLVRHE
TURTAFLVNDGVPYQLVGADGPGARVAAPTRVLSLLEPAEREAVRNLAART
PRPDAGALLRVVVARLADDDHVLVSTHIVSDAWSGVVLDELGRLEVCVTGPA
ALPPAVAGVQVAVQRAWAGVQBEHLAYWKRALDGSPVLRPFMDHPRPAVOSR
RETETEPALDVAALKEGROGATLFTLLAGFOVLLARHAGQEDIVVGVPAAGT
GTEFLGHPVLACADGVPGDAKRLTTLTAPFTFDSMBQLSMWLAGHELFIVPEDVR
QATFLGHPVLACADGVPGDAKRLTTLTAPFTFDSMBQLSMWLAGHELFIVPEDVR
RDSALVRVREHRIIDVITTSQELLSHGLLDGEMAPSMVVGGEAVSPSLMRTL
RDQRTRCNFLYGTATDATCDLSDPADVPVIGTPLPHTRVRLDRLRPVGV
AGSIYLGTLGALRFPALTARRFVADVPDTPGSRILYRTGDRMRPDPGTLEYLG
RTDQIKIRGRFVEPEIEAVLTHPANKVAAVDDAHARLVAVYVTLAEGGAGPTDV
RFAQGRPLPAHVPVSVAWVLEALPLTSNGLLDRALPAPAAGRPELDVRFVAPRDMVE
EVVAQWCAVLGVDRGVHDDPFELGSHSLVVQVMTIRKLLGVPEVPLRELFDAATV
EELARVRAARTEGLRGAAPPLGVDRSGLPLSFAQORLWLDOLA PDSVYNMCD
AYRAGPLDLDLURKALRTVERHETRLTAFVDRDGVPHQVVSAPPAPARAEEVVR
IEAAGRTDEARVLAABARTPRPDAGLMAVARLADDDHVLVSTHIVSDGWS
VDLDELGRLEHYHTGDPAGLPDVOVAVFVWQSWMTGSPVREHLAYWKRALD
GAPSLRLPADHPRPAVQSORGETVEFPLPAPLVARLEALCRQGVTLFMAIFGAFQV
LLARYSQDDVPLVPGVPTANTRAETPLVGFVNTPLVRVACSPSLFRALLDRVREA
ALGAFHQDLPLPFALEALAPEDLGHPLVOVTFOLLADPDERILVHGTDCVSLGFG
GVTSRFDLSLDVVGARGKCVITYCPDLFDRPRMEVLGHYTLTLAGAADDPGLRVG
DLPLSDDDVERLLGSGRPYLPAGAEITVPDAFAQVRAITPDAPALVHGDSTLTFAE
LDTRVLAALVLRRCGVAAETPVACLPRSDAVALLAVLRAGVYVFPVDPGSR
GAPGVVVRHGSVYLHLLGHVRMAEGPRNVAHTTAMTFDPSLEOLFVLAAGTILHV
VAPEVDRPALVALVRAAI DVLTNTPSHLTLLIAGLEGGDRVPGTPLVHVRVLDORL
ALMTRTERTGRTFNLGYTPTATVATCDLSDPADVPVIGTPLPHTRVRLDRLRPVGP
RPVVGAGVGEIYLGTLGALRFPALTARRFVADVPDTPGSRILYRTGDRMRPDPGT
GTLEYLGRDQIKIRGRFVEPEIEAVLTHPANKVAAVDDAHARLVAVYVTLAEGGAGPTDV
PRAPHGSDAGDPAQVEEWNVAETHTDADGELTFNKGWNSDLSLTKAPIPAEHR
EWDVTVARLERPAERVLIEIGSTGLLMWRLPHVTEYGTDFSRPAVMDLRDGLR
RPAHRLVLRHREATDFGVRAASTDLVWVSVQYTPDRAYLDTVLARALDADAGR
VFVGDVRLALAQFYAROLAHAGPCAAARDVARAAGEFAAMDGELLVSPAFVLAAL
ARSPRVGEIIFRRGRHNMESLYRDVVLVHVGDRPAPAEAVLTWGDQVDFDLASL
SARLGGPDPALLVRGANDRLDNELLPARTTAVEPELDWGLADSTPYRVSWSW
AAADPRGAMDVLLVRDADDDGLLVPHPPEPSAPLNTPTPHPSARQGSAAAGLR
SWLAERLPAHLIPARITEVDALPRTGTGLDRGALGLVTAGRGARAGDRPATAPRTG
LERTADAMARVLGPEVGVHENFVALGGDSLLAVRAVARCRAGVRLTVRQLLSQT
VAALAAALEESQ"

CDS

complement (18307..18660)

/codon_start=1

Alignment Scores:

Pred. No.:	8.11e-05	Length:	77457
Score:	566.00	Matches:	681
Percent Similarity:	30.64%	Conservative:	275
Best Local Similarity:	21.83%	Mismatches:	1098
Query Match:	5.04%	Indels:	1078
DB:	1	Gaps:	156

US-09-522-753-5 (1-2517) x AF210249 (1-77457)

Qy	3	GlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrProProHis	22
Db	64608	GGTGGCGAGCGCTCCCGCTGGCTGCTTCGAGCTACGACCA-----CCGCTTGA	64661
Qy	23	SerLeuSerTyrProVal-----GlnIleAlaArgThrHisThrAspValGlyLeuLeu	40
Db	64662	CGAGTCCGCGACCGCTGACCTGAGCTTGGACCTGCTGCTAT-----	64703
Qy	41	GlutyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleGln	60

Db	64704	-----CGCCGACGCCCCACAGCGTCCACGTACTACCGCGGACCTGCTCACCCTTCTACGC	64757
Qy	61	ProGlnArgArgProSer-----LeuLeuSerGluPheGlnProGlyAsn	76
Db	64758	CGACCCCGCGCGCCCTGCCGCCCTCGGTGCTCTCTCCGAGCTAGCTCTCGGCGT	64817
Qy	77	GluArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGly	96
Db	64818	CGCGC-----CCAGCCGCGAGG	64835
Qy	97	LysSerGluMetGluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspPro	116
Db	64836	CGA-----GCCGCGCGCGCGC-----CCTCGACCA	64862
Qy	117	LeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThr	136
Db	64863	CGCGCGCGCGCGCT-----GCCACCTGCGCG-----	64892
Qy	137	LysAspArgSerLeuThrGlyLysLeuGluProValSerProProSer-----	152
Db	64893	-----CCGCGCGCGCTCGCGTCCGCTCGCGTCCGCGCGC	64922
Qy	153	-----ProProHisThrAspProGluLeuGluLeuVal	163
Db	64923	CGAGGAGCTGACCGCGCGCGCTTCCGCGCTCACCCCGGACTCG-----	64970
Qy	164	ProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGluLeu	183
Db	64971	CCCCGCGCGCTCGGCGCGCTCG-----GCCGCGCGCGCGCGCGCGCGACT	65018
Qy	184	ThrMetValGluGlnIleSerLysLeuLysLysGlnGlnGlnLeuGluGlu	203
Db	65019	CACCCCGCGCGCTGATCTGCGCGCTTCTGCGCGCTCTCGCCAGCTGAGCGGACAC	65078
Qy	204	AlaAlaLysProGluProGlu-----LysProValSerProProIleGluSer	221
Db	65079	CCCCCGCTTCCACCTCACCTCACCTTCCACCGCGCGCGCTGCTCCCGCGGTGA	65138
Qy	222	LysHisArgSerLeuValGlnIleLeuTyrAspGluAsnArgLysLysAlaGluAla	241
Db	65139	CGACTCTGCGGACTTAC-----CACACGACCTGCTCGGGTCGACGGCGGCGGA	65195
Qy	242	HisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSer	261
Db	65196	CAC-----CTTCCGCGCGCGCGCGCGCTTCCAGGACCGCATCTGGGAGGACCTCGA	65249
Qy	262	AspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIle	281
Db	65250	ACACCGCGCTGTCAGCGCGCTGAGGTCTCGGATGCTGCG-----	65291
Qy	282	LeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArg	301
Db	65292	-----CCGCGCGCGCGCGCACCCACGACG-----CGT	65318
Qy	302	TyrAsp-----GlnLeuMetGluAlaLeuGluLysValGluArgIleGluAsnAsn	319
Db	65319	CGGATGCGCGTCTTCTTACCAGCACCTTCCGCGCGCGCGCGCGCGCGCGCGCGCG	65378
Qy	320	ProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGlu	339
Db	65379	CCGCGCGCGCTGCGCGGTACGCGCGCGGTACGCGGTACGCGACCGCGCGCGCTCT---	65435
Qy	340	IleArgLysGlnArgGluLeuGlnArgMetGlnSerArgValGlyGlnArgGlySer	359
Db	65436	-----GCTCGACCATCAGGTGAGCGAGAGCGACGCGCGACT	65471
Qy	360	GlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleLeuAspGlyLeu	379
Db	65472	GGTCTGACCTGCGACTACGCTCGCGGACG-----	65501
Qy	380	SerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProMetLeu	399

QY 922 SerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArg-LeuLeuSerProAr 941
 Db 67424 -----GGTGTCTGGGAATCGCGCGCTCGAAACCCGA 67453
 QY 941 gProSerLeuLeu-ThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProL 961
 Db 67454 TATGAACCTCTCGAGCGCGGTG---CCACCTCCGTCGAACCTCGTCCGCTCGGACCCG 67510
 QY 961 euAspLeuLysGlnLeuLys-----GlnArgAlaAlaAlaIleProProI 976
 Db 67511 TCTG-GAGGAGAACTCGGCTCGACACCGACATCGAGAACTGTCTGCTCTCCCGTCGG 67569
 QY 976 legInValThr-----LysValHisGluProProArgGluAspAlaA 990
 Db 67570 TCGCGGTATCGTCCGCGCGCACTCGCGCGCGGACCGACACCGCGCGCGGACCCCG 67629
 QY 990 laPro-----ThrLysProAlaProAlaA----- 998
 Db 67630 TCGCGCGCGGTCTGGTAGCTTTCGACCCCGGTCCGTACTGCGCGCGCGCGCGCGCG 67689
 QY 999 -----ProProProGlnAsnLeuGlnProGluSerAsp- 1010
 Db 67690 GACCCGTCGCGCGCGTCTCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67749
 QY 1011 -----AlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProA 1027
 Db 67750 CCTCACCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67797
 QY 1027 laProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuLeuGlyAspProP 1047
 Db 67798 TCCCGCGCGCTCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67857
 QY 1047 roCysTrpThrSerGlyLeuProPheProValProProArgGluValLysAlaSerP 1067
 Db 67858 CCGCA-----CCCATCCCGCGCGCGCTC-CGT-----GCC 67886
 QY 1067 roHisAlaProAspProSerAlaPheSerTyr----- 1077
 Db 67887 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67946
 QY 1078 -----AlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuP 1097
 Db 67947 GACGCCACACCGCATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 68006
 QY 1097 ro-----ArgProPro-----ThrIleSerAnProProProLeuLysSerSerAlaLysH 1114
 Db 68007 CCGGACGACACCACTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 68066
 QY 1114 isProSer-ValLeuGluArgGlnIleGlyAlaIle----- 1125
 Db 68067 ACGTTCGCGGACCTGGCGCGCTCTCGGCGCGCTCTCGCGCGCGCGCGCGCGCGCG 68126
 QY 1126 -----SerGlnGlyMetSerValGlnLeuHisValProTyrSer 1138
 Db 68127 GAACCAATAGCGCTATCCGTCGCGCGGTCTCTCTC-----TACCGCGCTCCAG 68174
 QY 1139 GluHisAlaLysAlaProValGlyProValThrMetGlyLeuPro----- 1153
 Db 68175 ACCTACCTGCTCTCCACCGCGGGAAGTGTAC-----GAGCTGCCCGCGCGCGCGCG 68231
 QY 1154 -----LeuProMetAspProLysLysLeuAlaProPheSer 1165
 Db 68232 GTCCACCGCGCGCAACCGCTGTGTAGCATCGACCCCGCGCGCGCGCGCGCGCGCG 68291
 QY 1166 Gly-VallLysGlnGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuG 1185
 Db 68292 GGCACCGCGAGATCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 68348
 QY 1185 yValProThrAlaGlnGlnAlaSerValLeuArgGlyThrAlaLeuGlySerValProG 1205
 Db 68349 ATCGCGCGCATCGACCGCATACAC-----CGCTCTACGGGATCTCT-----CCTGG 68396
 QY 1205 ygIlySerIleThrLysGlyIleProSerThrArgVal----- 1217

Db 68397 GACTTCA-----CGTCTTCGAGCGCGTGCATGACCCAGTTCTGCTGATG 68441
 QY 1218 -----ProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAs 1236
 Db 68442 CGGACCGCGCTCGGACCGGATCGGCTGTGCGCGTTCGACAGATGAGACCGCGCGCG 68501
 QY 1236 pValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAs 1256
 Db 68502 CTGCGCGCGGTTCGCTCCTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 68561
 QY 1256 p-----ArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrGlu-- 1271
 Db 68562 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 68621
 QY 1272 -----GlyLysGlyHisValLeuSerTyrGluGlyG 1283
 Db 68622 TCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 68681
 QY 1283 yMetSerValThrGlnCysSerLysGluAspGlyArgSerSerGly----- 1299
 Db 68682 T-----GGCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 68723
 QY 1300 -----ProPro-----HisGluThrAlaAlaProLysArgThrTyrAspMetMetG 1315
 Db 68724 GCACCCACCTCGACACCGCTCACCGCTGCTGCACCGTCAACCGAGC----- 68769
 QY 1315 uGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaI 1335
 Db 68770 -----CGTCTGT 68777
 QY 1335 eProProGluArgHisSerProHisLysLysGluGlnHisIleArgGlySerI 1355
 Db 68778 TCGCGCGCTTCGCGCTCGCGCTCGCGCGCTTCGCGCGCGCTTCGCGCGCGCTTCGCGCG 68837
 QY 1355 eThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAl 1375
 Db 68838 CCGCGGAGCGTACGTGCAACCGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 68895
 QY 1375 aLysLeuLeuLysArgGlu-----GlyThrProProProProProProSerArgAs 1392
 Db 68896 ATCACCCTTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 68945
 QY 1392 pLeu-----ThrGluAlaTyrLysThrGlnAlaLeu----- 1402
 Db 68946 ACTCTTACCAACTCCGCTTCCCGGACGCTGTCTTCTCGCTGCTCACCCTCGCGCGCG 69005
 QY 1403 -----GlyPro----- 1404
 Db 69006 CCGCGCGCGCGACCCCGCTCGCGCGACACCGCGCGCGCTCTGGCGCGCTCGACCGCGCG 69065
 QY 1405 -----LeuLysLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyAr 1423
 Db 69066 TCGTCCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 69125
 QY 1423 gSerIleHisGluIle-----ProArgGluGluLeuArgHisThrPr 1437
 Db 69126 G-----CATGCTCTGCGACGCGCTTCCAGACCGCGCGCGCGCGCGCGCGCGCGCG 69178
 QY 1437 oGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGly----- 1454
 Db 69179 TACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 69235
 QY 1455 -----ThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysHis 1471
 Db 69236 ACCGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 69295
 QY 1471 sAspValArgSerLeuIleGlySerProGly-----ArgThrPheProProValHisPr 1489
 Db 69296 CACGCGCGCTTCTTCCACGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 69355
 QY 1489 oLeuAspValMetAlaAspAlaArgAlaLeuGlu----- 1500

Db 69356 CAGTTCGACGAACCGGACCTCCGAGCGACCTCTGCTACGGCGAGCGCGCGCCCATCGAA 69415
Qy 1501 -----ArgAlaCysTyrGluIuSerLeuLysSer----- 1510
Db 69416 CCGCGCTCATGGAGAACTGCACACCGCTACCGCGCGGAACTGGTGGCGCCCGCTGG 69475
Qy 1511 ArgProGlyThrAlaSerSerGlyGlySerIleAlaArgGlyAla----- 1526
Db 69476 CGGCGCGGAGCTCTCTCGTCGACAACTCTCTACCGCGCACGGCAGGAAACCTTTC 69535
Qy 1526 ----- 1526
Db 69536 ACCGGGAAACCGCGCTGCTCGCGCATGCGACAGCGCTGGACTGGAGCGAGGTGAGC 69595
Qy 1527 -----ProValIleVa 1530
Db 69596 GCGTGACGCGCGCGGCACACCGCTGCGCGGACCTTCGTCCAGCGCGCTGTGGCGGT 69655
Qy 1530 l-----ProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAs 1545
Db 69656 CCACTGCTCACCGCGCGCGGAGGTCAACCACTCTCTACCGCGCACGGCAGGAAACCTTTC 69715
Qy 1545 pHis-----GlyAlaProPheAlaGlyHis----- 1553
Db 69716 ACACGACACGCGCGCTCAACGAGCGCTCCG-----GCGGGTACCGCGCGCTCCCG 69772
Qy 1554 -----LeuPr 1555
Db 69773 CCTCACGCGCGAACTCTCGCGCAGGAGAACCCCGCTGACCTCGCGCGGACGCGCC 69832
Qy 1555 oArgGlySerPro---ValThrMetArgGluPro-----ThrProArgLeuGlnI 1571
Db 69833 CCGAGGTCAACCGCGTGCACCTGCGCGGAGCGCGCTCGCGGACGCGACCGCTCTGCG 69892
Qy 1571 uGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgG 1591
Db 69893 TGGCGCTGCTGCGCGCGCAGCGGACCACTCTCGCGCGGCGGACCGCGCGCGCTTCC 69952
Qy 1591 uIleAlaLysSerPro-----HisSerThrValProGluHisProHisProIleSe 1609
Db 69953 ACCTGGTGGCTCCACGACGACGAGCGGTGCTCGCGCTCACGCGCCACACCT- 70007
Qy 1609 rProTyrGluHisLeuLeu-----ArgGlyValSerGlyValAspLe 1623
Db 70008 -CTCTCTCGACACACCGTCTCTCTACGCGGTGCTCGCGCGGTCTGCGCGGTA----- 70061
Qy 1623 uTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLe 1643
Db 70062 -CGCGCGCGCTTCCG-----CCCGAGCACTACCGCA----- 70094
Qy 1643 uAspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnPro-----Th 1660
Db 70095 -----CGCACACCTTCGCGGCGGCGCGCGCGCGCGCGCTCTCCCGTTC 70138
Qy 1660 rTyrProHisLeuTyrProTyrTyrLeuIleArgGlyTyrProAspThrAlaLeu-- 1679
Db 70139 GGGCGCGGCTC---CGCGCGCGCTGGTGGCGCACCGCGCTGCGCGCGCGCGCGCG 70195
Qy 1680 -----GluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnMetHi 1696
Db 70196 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 70255
Qy 1696 sHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProAr 1716
Db 70256 TCCCGGACGCGCTGGAAGCCCTGACCGCGCTGACGCG---CCTGGCGCGCGCGCTCG 70312
Qy 1716 gGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerG 1736
Db 70313 GCGCAACGCGTCTCGCTCGCGTCTATGCGCTGCGCGC----- 70349
Qy 1736 nValProHisLeuProValLeuValProThrProGlyThrProAlaThrAlaMetAs 1756
Db 70350 -----CTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 70385

Qy 1756 pArgLeuAlaTyrLeuProThrAlaProGlnProPhe---SerSerArgHisSerSerSe 1775
Db 70386 -----CCGCTTACCACCGCTCTCGACCTGCGCGC 70414
Qy 1775 rProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrSerSerSerG 1795
Db 70415 ACCACC-TCCGACTTCGGCGCGCGCGCTCGCGCGCTTCAACCGACCGCTCTTTCGGCG 70473
Qy 1795 uArgGluArgAspArgAspArgGluArgAspArgGluArgGluLysSerIleLe 1815
Db 70474 GACCTCGCGGAGCGC-----CGCGCGCGCTCC 70500
Qy 1815 uThrSerThrThrValGluHisAlaPro---IleTyrArgProGly---ThrGluG 1833
Db 70501 TTCGGGACGCTCACGCTGCGCGCGCGCTTCTGGAGCGCGCTGCTGCACCTACCTC 70560
Qy 1833 nSerSerGlySerSerSerGlyGlyGlyGlyGlySerSerSerSerArgProAlaSe 1853
Db 70561 CCTACGCGCGCGCTGCGAACTCGGAGGAACTGGGCGCGCTCACCGCGCGCGCA-- 70618
Qy 1853 rHisSerHisAlaHisGlnHisSerProIleSerProArgThrGln-----As 1869
Db 70619 -----CCGCGCGCGCACTCGGAGCTGCGCGCTGAGC 70647
Qy 1869 pAlaLeuGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAl 1889
Db 70648 TTCTGCGCGCAACCGCGCA----- 70666
Qy 1889 aValGluProSerLysProThrValLeuArgSerThrSer-----Th 1903
Db 70667 -----CCAGCGCGCGCACCGCGCGGAAACCGACCTCGCGGAAACCGCGCTGTGCATC 70719
Qy 1903 rSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyG 1923
Db 70720 GAGTGTTCGCGAGGCGCGACCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 70779
Qy 1923 YThrLeuAspGlyValTyrProThrLeuMetGluProValLeuProLysGluAlaPr 1943
Db 70780 GACGG--CACGGTGTCTCGCCT-----CTCCCTAGCGGAACTCG 70816
Qy 1943 oArgValAlaArgPro-----GluArgProArgAlaAspThrGlyHisAlaPheLe 1960
Db 70817 GCGACGACACCGTGTGCTGCTCGACGCGCGACCGCGA----- 70853
Qy 1960 uAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerG 1980
Db 70854 -----CCACCGCGCACCGGAAACCGCGCGCTGCTCCACCGATGACGACCGCGCG 70906
Qy 1980 uPro-----ArgProLeuValProProValSerGlyHisAlaThrIleAlaAr 1996
Db 70907 TCCTGGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 70954
Qy 1996 gThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaProAl 2016
Db 70955 CCGCGCGCGCA-----CACACGAGGAGCACCGGATGACGACCGCGCGCG 70999
Qy 2016 aSerAlaSerAspPro-----HisArgGluLysThrGlnSerLysProPheSerIle-- 2033
Db 71000 GACCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 71059
Qy 2034 -----GlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPr 2050
Db 71060 GCTCGCGCGCAACGAGGATCCCGCG-CGGCTGGCGGCG---CACCGGAACTCCCGCACCC 71115
Qy 2050 oGluGlyValGluProValSerProValSerProSerLeuThrHisAspLysGlyLe 2070
Db 71116 AGGAGGAGTG-----CCTCGCGCACATCGACGAGGTGT 71148
Qy 2070 uProLysHisLeu---GluGluLeuAspLysSerHisLeuGluGlyGluLeuArg----- 2087
Db 71149 GGACCGACATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAGC 71208

QY	2088	-ProLysGlnProGlyProValLysLeuGlyGlyGluAlaHisLeuProHisLeuA	2107	Db	72195	TGCCCCGTCGCTCGTCCGCTCCACCCCGCGGGCACCAGTGGACAGGATGGC	72254
Db	71209	CCGCTCCCGCCCGCCCGCCGA---GGAGAGCCGAGCCCTCGTCGACCCGCTCGG	71265	QY	2389	ArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyAlaLysValSerGlyArg	2408
QY	2107	gProLeuProGlu-----SerGlnProSerSerProLeuLeuGlnThrAlaPr	2124	Db	72255	CGCTCGAGC-----ACTACGCCCGGCTCTCGCCCGGCTCGGGCCGAACTCTCGG	72305
Db	71266	CGGGGACACCGCGTGGAGTGGTCTCGCCCGGAGCGACCGCGCGCCCTCGGG	71325	QY	2409	ProSerSer-----ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAs	2425
QY	2124	oGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisLeuSerGluValleTh	2144	Db	72306	CTTCCTCTGCTGCTGGCGACGAGCGAAGCCCGAGCTGGAACCTGCTCGGGCAGC	72365
Db	71326	AGCCGTCGACCGGGCTAGTCTCTGTCGCTTCGCGCACCCCGCGCGGACCGAAC	71385	QY	2425	paRgProProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLe	2445
QY	2144	rGln-----AspTyrThrArgHis-----	2150	Db	72366	CGTCTCTGAGCGGCTCCCGCG-----AGCCGCTGCTCCACCTCG	72404
Db	71386	TCGGCGTTCGCTCGACCCCGCGCGACCACTATGAGCGCACCGAGCTGGCGCTGACCG	71445	QY	2445	uThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrPro-----	2461
QY	2151	-----HisProGlnGlnLeuSerAlaProLeuProAlaPro-----Le	2163	Db	72405	AGCGGGCGCGG-ACCTCGACCGGACCGAGACCTCTCGCGCAGCGCAGCTGCTGTC	72463
Db	71446	GCACCTCACCCTCGACTTCGACCGGTCTCGTCCGACGCGCGCTCGAGTGCACCACT	71505	QY	2462	-----PheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerPr	2479
QY	2163	uTyrSerPhe-----ProGlyAlaSerCysProValLeuAspLeuArgProPro--	2180	Db	72464	GCAACGACTCTCTCGCTCGCC-----ACGTC	72490
Db	71506	TCACGGCGAGGGCGCTCGAGCGG-TGTTCGGCACCTGACCCCGCGCGCACCCGG	71564	QY	2479	oProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisAlaTrpAs	2499
QY	2181	-----SerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerPr	2198	Db	72491	GC CGCGCGCTCGCACCCCGCTCGCTCTCTACGGCC-----	72530
Db	71565	CCGTGAGCGCGCTCGGACCGCGCGCCGACCCAC-----CGAAGGAGGAGACC	71615	QY	2499	pGluGluProLysProLeuLeuCysSerGlnTyrGluThr	2512
QY	2198	oHisSerGlu-----	2201	Db	72531	----CGACCGGACCGAGTACTGTGGACCGAGTACTACC	72566
Db	71616	CCATGACCCCATGACACCCCGACCGACCCCGACCCCGCGCGTCT	71675	RESULT 88			
QY	2202	-----GlyCysLysArgSerProGluPr	2209	AE007093			
Db	71676	TCGCCCACTCGCGCGCCCGCTCGCGGACCTCTCCAGCGAACATCGGCTCGCCC	71735	LOCUS			
QY	2209	oAsnLysThrSerValLeuGly---GlyGlyGluAspGlyIleGlu-----ProValse	2226	DEFINITION	AE007093	14860 bp DNA linear BCT 27-APR-2001	
Db	71736	TCGTCCGCGCGCGCGCGCGGCGGCGGTCACCTGTGCTCGCGGAGACTGCGGG	71795	MYCOBACTERIUM TUBERCULOSIS CDC1551			
QY	2226	rProProGlyMetThrGluProGlyHisSerArg-SerAlaValTyrProLeuLeuT	2246	complete genome.			
Db	71796	CCCGCTCGCTCGGCACTCACCCGCGCACAGTACCGCACCGCTGCTGCCCTGC---	71851	AE007093 AE000516			
QY	2246	yrA-gAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrS	2266	AE007093.1 GI:138822291			
Db	71852	-----CCCCAGCGG---GGCAGGCGCGACCCC	71875	MycoBacterium tuberculosis CDC1551			
QY	2266	erGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMet-----ValL	2284	MycoBacterium tuberculosis CDC1551			
Db	71876	--CGTGGCGCGCTCTCTGCGCACCTTGGCGGACCGCGGCTTGCCTCGCCCTCGCTCG	71933	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
QY	2284	ysrLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArg-----AsnGluP	2302	Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium			
Db	71934	ACCCGACAGCAGCGGCTCGACGCC-----GGCCACCGCGCGCGCGCGCGTGC	71984	tuberculosis complex.			
QY	2302	roGlu-----TyrAsnLysSerGlnProGlyThrGluLysPheAsnMetPro---	2317	1 (bases 1 to 14860)			
Db	71985	CCGAGCGGATCGGCTCGCGAGGACCGCGCGGAGACACATCACCATCCATCCATCC	72044	Pfeischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,			
QY	2318	-----AlaIleThrGlyThr---GlyLeuMetThrTyrArgSerGlnAlaValG	2333	Peterson, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,			
Db	72045	GCCTCCAGCTCCCTGTGGGGAGCCCGGACCTGTACGAGTACGCCATGCC-----	72097	Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,			
QY	2333	lnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleArgLysAlaLeuMetGlyL	2353	Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,			
Db	72098	-----CTCGCGCGCGCGCTGGGCTGCTCGCGCACCGCGCGCGCGCGCGC-----	72139	Gill, J., Mikula, A. and Bishai, W.			
QY	2353	ysTyrAspGlnTrpGluGlu-SerProProLeuSerAlaAsnAlaPheAsn-----	2369	Whole genome comparison of Mycobacterium tuberculosis clinical and			
Db	72140	-----GACGTCTCGCGGAGTGGCCCGACCCCGCGGCTGCTCGCGCGCGCGGTC	72194	laboratory strains			
QY	2370	---ProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGly	2389	Unpublished			

```

/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, LuxR family"
/protein_id="AAK46866.1"
/db_xref="GI:13882292"
/translation="MLASMSKIHPVDVVPVDSADGVSELVPTGVTLLADIENAT
HLPSQDQTIAKLDRITLTVREHGVCFVEQEGDSFLVAFARASDAVACALGIQ
RAPLAPRLIRIGMTGVSSPCNCGVGTIDRTARLRELAGGQTVLSTGTVLADVLAD
LLPKDAMLNDLGTVRLDPLRPVVLCHPDLHNAFPPLRTRKVVGAHCLPAQLTRL
VGRVDEVAQVRLGLDKRWTLTTCVGVGKTRLATOVASADVDPGVVYVNLAPIT
DPALVFAARVRLGPQGRSTVDTIVRIGRBRMLVLDNCEHLLDGCALIVALL
GACPALRVATSRPIAVAGEQIWRVPLGHGAIEULTDRAREARPELSEITADNLAL
VTIECHRLDGLPIALIASRVRALALTEIVDSLHDFRLLTGGSRIVARRQQTMRAS
VDMSHALLTGPQVLFRLAVFPFGFDLGAQAAAGDVQRYEVVDLLSLLADKSLV
LDDSDGRTYRLLETVQVLALEKRESGDADAVRARHRYAAVAAAGLDPVSAVAGE
RLNQAREITDNLRAAFPSRENGDTHALLASCILOPLWARGRLOEGLWFAAALA
DHDAHPAGDPGLYARALADRALIDAVAGITDRLLDQAQALATARDIEDPALLARALT
ACGVAAYNADLAPLAEAVGLARAVGDKRWLAELVLAQVYVGFAGEGDPGATRAAG
BEARSLADEIGDAPLSRSCRWALAAANLWQNLAEAVGLSREVIAGESDAADHVVSCA
GOACLAHALAHRGDTAAQAAQASIDTAVGLSPVLSGACSAIVFATLAAGDVAAAEH
ARESATRFPCASAAAIINDPTSSAQISCARGDLNAAHRLADGAASITRQVHRAALTT
RCRIEIAQGRHRAERDAHDALGVAASIGAYLWPDILLECLASVMADAGSNREARLF
GADARGRMGAVRFGIYQAGCNSSLATLRKSMGDSFDDAWAEGTALSIDEALYAYQ
RGRGARKRTPSGMGALPTTLEVALLVLEGLSNKKGIVRLFISPTVHSHLTHVYTKL
GLSSRLQAQAAARRGESRGPSP"
/complement(3451..3813)
/gene="MT2563.1"
/complement(3451..3813)
/gene="MT2563.1"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46867.1"
/db_xref="GI:13882293"
/translation="MAPPRWALURGRATLGSFSAWLGVTAKAAEAAAPSSFPFLRKP
HRAGDSADRAGDPDGTADHADVSVLAGDAASTGGLTASQGHGCHRSAAAMARRSPNA
STKARTRHGPAAKRFRAI"
/complement(3859..8856)
/gene="MT2564"
/complement(3859..8856)
/gene="MT2564"
/note="similar to GB:AL123456; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="PF_PGRS family protein"
/protein_id="AAK46868.1"
/db_xref="GI:13882294"
/translation="MGRCEMSYVIATPEMMATAAFDLARIGQVSAASAFAVAMPPTTEV
VAAGADRVNAGIAALPSAHAQEQALSAQAAAFHDFVHTLTAAARWYTATETANAA
MRVVLGVNATPTQLLGRPLIGDAHCTAPQPGGAGLLFNGNGNCAAGVQVGA
GGAALFPGICGAGAGAPPGTGTGTGWLACGGVGMGGAGGGAGGAGNAGLFG
NGAGAGGAGGAGGAGGAGVGMGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
GAGDGLAGDGDGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
ATNGEPAGI SAAGDGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
GLGAGPCDAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
AAGNDGPVSGGKGGNGAHAPVAGHGGNGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
YADMTLIFLSSGTPGIDGNGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
GFNAVLVSDGNGDGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
NGSGAGRVDFKDGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
KCGNADATVAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
GQPEGSEANGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
GVGGDGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
GDPFGAGLGLGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
GDGNGRVDGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
GDGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
VAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
AAAGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
GDDGGKGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG

```

```

GSPGGGGVGGHGGAGGAGAGMNGGGGTGGGGGNGAAGGAGWSPPDULKGFDFDGGG
GAGCGGAGGAGGTQTDGDDGGAGGLGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
YGGWGGAGGNGGAGGAGGAPAGVGNRGGVGGGGGGGAGGAGGAGGAGGAGGAGGAGG
EPGCGDGGDGGGWRPQGGGFGSTSGAHGKAGFAGPGAGGAGGAGGAGGAGGAGGAGG
ADAGDGGGNGGNGGLGAGRGDGGAPGGDGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
GRGDDGRGAGGLTSGVSGSGSGSGNGRGDPGSGSGGEGGEGGEGGEGGEGGEGGEGG
complement(8928..9134)
/gene="MT2565"
/complement(8928..9134)
/gene="MT2565"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46869.1"
/db_xref="GI:13882295"
/translation="MPAGSLMAGMBEIDPGADVAPLDCKSVKDDVGNPVAAGSVALL
LADVSTHLGRRGKSEQLSRR"
/9270..9893
/gene="MT2566"
/9270..9893
/gene="MT2566"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK46870.1"
/db_xref="GI:13882296"
/translation="MVDTSAPASRLDTPRAHVSLSKHPYQIQVFGSGGTIGPRVVEL
AYQVGAETAKQGHILISGTTWEASRGASDADGLVGLPDKFTDGNASTIKI
LSQMFPARNYITGLSCHGAIVVGSGSAYEERVRVWEGRPVVLANSGSPGTGASQOM
LSMQEIFYVAPFDPKPKWRVFSATPAESVSLVIGLIRKQAQHEP"
/9883..10635
/gene="MT2567"
/9883..10635
/gene="MT2567"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46871.1"
/db_xref="GI:13882297"
/translation="MSRIINEFGVIYATIGTWAGLVRVLDLGSQCQFDEDRERI
ALNSVIRKSSVQNYPDLTIEHCNSAQLKAMLDPMFTMDIDVVVKSFSRGAHSVH
RRIKEGRMIEFVIERLSLIPESKAVVVFPTTYEDYAAVMNRHDDYLPCLVSIQFRLL
PDGKDYVHTFYRSRMDAMQKGNLLSTAKLSDWVRENVASIRIGRKIMLPLDGM1
CDVHIYKTYAEACKRLANLDRRTQFDVARN"
/10688..10909
/gene="MT2568"
/10688..10909
/gene="MT2568"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46872.1"
/db_xref="GI:13882298"
/translation="MRTTLDLDDVIAAARELASSQRRSLGVSISELARRGLMPGRVE
ADGLPVIRVPAGTPTPTMVRALDE"
/10915..11340
/gene="MT2569"
/10915..11340
/gene="MT2569"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46873.1"
/db_xref="GI:13882299"
/translation="MALLDVNLVALWDSHIIHARTREWFATANATIGMATCPITEAG
FVRVSTNPKVLPISAIGIADARPVLVALRAGVGRFLADDSLVLDVDDVPLVIGRVQVTD
AHLTLARRRGVRLVTFDAGVFTLAQORPKTPVELLITL"
complement(11363..12544)
/gene

```


QY 1131 lGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetG1 1151
Db 4968 -----CACACGCCCGCGCGCCACCA-GCCCCACCGTGTCCCGCCAGCAACGGTGA 5019
QY 1151 yLeuPro---LeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnG1 1170
Db 5020 CCCTCCGAGGACCGTGGCGCGCGTCCGACCCCTTGGCGCGCTCA----- 5065
QY 1170 uGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaG1 1190
Db 5066 -CGCGGTACCGATGTGGCGCGGTACCGGCC-----TTGCTTCCAGCCCG 5112
QY 1190 nGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLy 1210
Db 5113 ACCGGCCCC-GCCATACCGCCGAGAGCTTCGCGAGCGGTGGCGTCCGTCGCCATCACAC 5171
QY 1210 sGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleTh 1230
Db 5172 CGCTCCGCGTCCCGATAGCCCGCGTGGCGCGTCCGCGTCCGCGCGCGTCCG 5231
QY 1230 rHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyLys 1250
Db 5232 CGCGCGCGCGCGCGC-----CTTGGCACCGTTGCC----- 5262
QY 1250 pSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTy 1270
Db 5263 -GCCGACCGCGTCCCGCGTCCGCGTTCAGTCCCGCGTGGTGGCGTCCACCGCGT-- 5319
QY 1270 rGluGlyLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCys 1290
Db 5320 ----GGCGCGGAGGTGCGTGGCGGTGTCACCCCGACCGCGCGCGCGTGGCG 5375
QY 1290 rLysGluAspGlyArgSerSerGlyProProHisGluThrAlaAlaProLysArgTh 1310
Db 5376 C-----GGGTCACCGCGGTATTACCGCGCGCTC 5402
QY 1310 rTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLe 1330
Db 5403 CGCGCGCGCGCGCGCGTATACCAATGCGCGCGT-----GCCGCGCGCGCGCGCAACGC 5459
QY 1330 uMetGlyArgAlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHisH1 1350
Db 5460 CGCGCGCGCGCGCA-GTCCGCGCGTCCGCCACCGACCCCGCGTGGTACTGCCGA-- 5516
QY 1350 sIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnAspTy 1370
Db 5517 -----CCGTCCCGAAGGTGCGCGCGCGCGCGCGCGCA 5548
QY 1370 rLeuArgArgGluAlaLysLeuLysArgGluGlyThrProProProProPro 1390
Db 5549 CGTCCCGCGCATGTCCACGACCCCGCGCGCACCG-CCATCGCGCGCACACCGCGCGC 5607
QY 1390 rArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLysProAl 1410
Db 5608 CCCACCGTGGCGCGCGTACTGTGATA---CCGTTGGCGCGCGT----- 5650
QY 1410 aHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProAr 1430
Db 5651 -----GCCCGCGC 5658
QY 1430 gGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGly-- 1449
Db 5659 CCCACCGTACGCGCCACACCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5714
QY 1450 -SerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLy 1469
Db 5715 CGCGGTACCGCACCGACCG-----ACCGCGCGCAC----- 5748
QY 1469 sLysHisAspValArgSerLeuIleGlySerProGlyArgThr-PheProProValHisP 1489
Db 5749 -----GGCACCCGCGCACCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5783

QY 1489 roLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuL 1509
Db 5784 CC----- 5785
QY 1509 ySerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValI 1529
Db 5786 --GCGTGGCGCTGACCCCGGTTCCTCCCGGTGGCGAGGCGAGGGCATCT- 5836
QY 1529 leValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaP 1549
Db 5837 -----GAACACGAGCGCGCGCGTGGCGCGCGTGGCGCGCGCGCGCATCGCCAC 5888
QY 1549 roPheAlaGlyHisLeuProArgGlySerProValThrMetArg-----GluProT 1566
Db 5889 CA-----TTGCGCGCGCGCGCACCATCGCGAGCGCGCGCGCATTCGCGCGCGT 5936
QY 1566 hrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuT 1586
Db 5937 CGCTCCCGTTGCC-----C 5951
QY 1586 hrSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHisHisProH 1606
Db 5952 CGCGCGCGCGCGCGCGTGGCGTTCGCGCACCGTGGCCACCGTACACCGTTCGCGCGCG 6011
QY 1606 isProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgS 1626
Db 6012 TGCGG--GGCGCGTGG-----TCGCTGGCGTCCACACCGCTG--GCCG 6050
QY 1626 eHisIleProLeuAlaPheAspProThrSerIlePro-ArgGlyIleProLeuAspAla 1645
Db 6051 CGCGGTGGCGTGGCGTTCGCGCGCGCGCAAG--CCGCGAGCGCTCCCGCGCGCGGGT 6107
QY 1646 AlaAlaAlaTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyr 1665
Db 6108 CACCGCGCTACCGCGTCCCGCGCATCA-CTGCCATGACCGCGCTACCGCGGTGGCG 6166
QY 1666 Pro-----ProTyrLeuLeuArgGlyTyrProAspThrAlaAlaLeuGluAsn 1681
Db 6167 CGCGTCCCGCGCGCGCG-GTCAACCGCGCGCGCGCGCGCGT----- 6216
QY 1682 ArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisThrAlaThr 1701
Db 6216 ----- 6216
QY 1702 AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAla 1721
Db 6217 -----GCCGACCTGACCGCGGTGGCGCGCGCGCGCGCGCGCATC--- 6258
QY 1722 LeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuPro 1741
Db 6259 -----ACGCGCGCGCGC----- 6270
QY 1742 valLeuVal-ProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLe 1761
Db 6271 --GCCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6307
QY 1761 uProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyG1 1781
Db 6308 -CTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6361
QY 1781 yProThrHisLeuThrLysProThrThrThrSerSerSerGluArgLysArgAspArgAs 1801
Db 6362 -----TTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6390
QY 1801 pArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThrVa 1821
Db 6391 -----ACCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6437
QY 1821 lGluHisAlaProIleTyrArgProGlyThrGluGlnSerSerGlySerSerGlySerSe 1841
Db 6438 CCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6470
QY 1841 rGlyGlyGlyGlySerSerSerArgProAlaSerHis----- 1854

RESULT 89
 APO05045/c
 LOCUS
 DEFINITION Streptomyces avermitilis genomic DNA, complete genome, section 25/30.
 ACCESSION APO05045
 VERSION APO05045.1
 KEYWORDS GI:29609643
 SOURCE
 ORGANISM Streptomyces avermitilis MA-4680
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
 1 Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C., Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T., Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
 Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites
 Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
 21477403
 11572948
 2 Ikeda,H., Ishikawa,J., Hanamoto,A., Shinose,M., Kikuchi,H., Shiba,T., Sakaki,Y., Hattori,M. and Omura,S.
 Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis
 Nat. Biotechnol. 21 (5), 526-531 (2003)
 22608306
 12692562
 3 (bases 1 to 299925)
 Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C., Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T., Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
 Direct Submission
 Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
 (E-mail:bioinfo.go.jp, URL:http://www.bio.nite.go.jp/, Tel:81-3-3481-1933, Fax:81-3-3481-8424)
 This work was done in collaboration with Haruo Ikeda(*1), Jun Ishikawa(*2), Akiharu Hanamoto(*3), Chigusa Takahashi(*3), Mayumi Shinose(*3), Hiroshi Horikawa(*4), Hidekazu Nakazawa(*4), Tomomi Osonoe(*4), Norihiro Kishida(*4), Hisashi Kikuchi(*4), Tadayoshi Shiba(*5), Yoshiyuki Sakaki(*6,*7), Masahira Hattori(*1,*7) and Satoshi Omura(*1,*3).
 Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa
 *1 Kitasato Institute for Life Sciences, Kitasato University
 *2 National Institute of Infectious Diseases
 *3 The Kitasato Institute
 *4 National Institute of Technology and Evaluation
 *5 School of Science, Kitasato University
 *6 Institute of Medical Science, University of Tokyo
 *7 RIKEN, Genomic Sciences Center
 Following url is also available.
 http://avermitilis.lis.kitasato-u.ac.jp.
 Location/Qualifiers
 1. 299925
 /organism="Streptomyces avermitilis MA-4680"
 /mol_type="genomic DNA"
 /strain="MA-4680"
 /db_xref="taxon:227982"
 /note="This strain is also named as strain: ATCC 31267, NCIMB 12804 or NRRL 8165.-synonym: Streptomyces avermectinius"
 140..1522
 /gene="amyA4"
 140..1522
 /gene="amyA4"
 /note="SAV5981"

/codon_start=1
 /transl_table=11
 /product="putative alpha-amylase"
 /protein_id="BAC73693.1"
 /db_xref="GI:29609644"
 /translation="MARKALAVSLAALAGLTVPAVSPPTTAAASPPTKDVAVLFEW
 KFDVSARECTNLGAGYVQVSPPAHIGSQWTSYQVSYKILAGRLGDAFAQN
 MINTCHTAGVKVVDVTVMHMSAGSGTGGSAITKYNPGLSYSSYMDCTATIDY
 TNRANVONCELVLGLADLTGEEYVKRTIAGYMTLLGAGDGFVDVAKHILPAADLAN
 IKSRLTPSVYKQEVYIYASGEAVQPTGYTGVQDQYERAYADLKRVPNNENLAYLKN
 YGEGWYLNNSVAGVVDNHDTERNGSTLANKGANTYLANVFLMPLAYPGAPDINSYG
 EWSADAGYPCGGTVMACQDWKQCHAWPEIKAMVAFNRNATRESVFTNWDNGDAIL
 AFGRGAGYVAIHESGLSTRTYQTSTAGTYCNVNNTGTVDSSGRFTATLGANTTA
 LALYSGKSTC"
 1661..7081
 /gene="amyX"
 1661..7081
 /gene="amyX"
 /note="SAV5982"
 /codon_start=1
 /transl_table=11
 /product="putative pullulanase"
 /protein_id="BAC73694.1"
 /db_xref="GI:29609645"
 /translation="MDLIPRGSAPRTRRTHSRSTAAVAALAAVSLAQPLEARAAT
 /PAPPSDLAAEPARHDAIREQYFVMPDRFANGDTSNDKGLTSGSLSTGYDPTDK
 GFYQGGDLGLTRKLDYIKGLTTSIMLAPIFKQPVQGTGKDGASGYLKGAFVYLTK
 VDHPGTWKDLTLISKAHGKMKVFVDITNHTADVDEKSYLSKSGAPVYLTK
 DGRPDDAGYTDGPKFPAVDGDSFRTPAVAARKNAKVPNSWLDPTMYNHRGSDTF
 AGESTHGDGSLDLDLTERPEVVRGMEKIYEKVVDRDGDIGDFRIDTIVKVNTERPQ
 WATALDAYAKRGKDDDFPMFGEVYSADTSVTSVYVQGRDSTLIDFPQDAARSVASQ
 CGSAAKLASVFGDDYKTTDKANAYEQVTFILGNHDMGRIGYFLNODNPKATDAELLRK
 DRLANLFLSRNPVYVYDEQGTGGGQDKARQTMFASKVADYLDLDDDEIGTRGH
 ASDVATSAPLYKEIALSKLRKNPALADGIOTERYAADGAVAFRTDARTSTEY
 VVAVNNAKASATFATGSADTAFKGIHTDDVLKSDAKKITVTPAGAAVILKAAG
 RGTTPAARSLTLKAPAGAGTVELSADVDGRLNRVVFPAQVAKNAKWTLLGSADHA
 PRYTPKGVDPAGTALRYKAVIDAAAGRTASATAASTGTGTPPAETPTASSRDYAI
 VHYKPDGDDYTDRLYAWDGLDGESTVPGHDFVDRDAGVAFVILKPGASVNF
 LIVDKDGQDSADRTIDVTKAGEVVMVQGETVTERPDYPAQDKTAVIHYHRADG
 DLTGMLHWITGAATPDWSPLEPVDAYGAVEPLTDGATGSLSYIHKDGSKDL
 SADSLDLTADGHEVWLLNGOENHLLPOPAGSAAALDLTTSKAVINDRWNTWNGSDA
 AASTQLSRDGSIAKVGSLTSDDERWLRSLKTSLTDAQKAAPHLKSGVATSDR
 DDVREALAGVQASQAANGAVLAATGVGVLQGLADLIDYDKADLGETFRGCHPTL
 AWMAATQSVSLELDGAVHVRMKNATGVSVTPASWKGKPYRVVVKVAVPTVRKV
 TNKVTDPYVALTIDTSERSLVVDLDRSLAPSGNSLKKPKAPLDRDAIEQLHTRDF
 SVADRTVPAKDRGTLYLAFTDKNSDGRSLROLAESGTSYVHLLPAFDATIAEKKSQ
 QATCDLASAADSEKQOECLTAAVAKDAYNMGYDPHYVTPEGSVATDANGTRTVE
 FRMWKSLNQDGLRVVMDVYVNHNTAAAGAGTSVLDLRIVPGYQRLADGVAATSCC
 ANATATENAMGKLVVDLSVTVWAKYKVDGFRDLMGHQPKANILAVRKALDALTAKD
 GVDGKKIILYGEWNFGEVADDARFVQATQKNMAGTATFSDRDARDVARGGPPDAD
 PGVQGFSGLYTDPNSDANGTPAPQKARLLHYODLIKVGLSGNLAKYRFTDSSGKEV
 TGSEVDYNGTGAGYADAPGDALAYADAHNLSLYDALTYKLPKGTGAGDRARMQVLAM
 ATAAALAGTSLQAGSDLLRSKSLDRNSYDSGDWENAIHWNQDNGFGRLGPMADIN
 AKSKPYATPLLSVVKVCGDQIEGTSAGYQDLIRITTEPESLSTAGOVQSKLTPPLS
 GKDETPGVITMKLGLDVLVFNFAVTPDQOQETVAALAKGVALHPVQAGADPIKVSAY
 TAKSGMFAVPGRTVAIFSQAR"
 7152..8756
 /note="SAV5983"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC73695.1"
 /db_xref="GI:29609646"
 /translation="MDANGROTDSTVWVDDVPTSRYAMGAVLRAGHSVVPVASAGE
 ALVELDLVLRAGSLPDVALVDGLPDMSGFELCRLLKQSPPTAALPVHVFSAALAPG
 DRQGLDAGDAYLSVPAPEPEIQAIVRAVARGARLTGAELARLRLTLLSETIVAVQ
 AKRCPTELAHAAQAGARTITDFAAVFLVGEGLSYPLGSRRAPELTLPGGAHEAVA
 RLRTGTAGAGVRSIVVPAPLPAFLPAGFRLPVSQEDARVLAQTEQKQKPAVCLATPAH
 RAQOTEIGVDFATLRSGVLTGVGVVHSLDAAATVMEIRHALRAYCEVDDPPA
 VLVTRDLRMLQHVHPEATATVCLALVPDPSGHVSIANAGHIPPLIVHDDGTATPGAA
 GPLGLGLRPPATEATVTPRHLRLMTDGLITETRGTDLSVSMQRLMAAADPGSGVG
 ALCDALLDCFGDRDDIALLALRLG"

CDS

FEATURES
 source

gene
 CDS

CDS	8788..9144 /note="SAV5984" /codon_start=1 /transl_table=11 /product="putative isomerase" /protein_id="BAC73696.1" /db_xref="GI:29609647" /translation="MPQITVDYSAEALAPDFRPGFARLHEEVVTRTAAAKEPEAKTQF RGEEDTVVGDGEGHAIHVHTLGLLAGRSDETKARLTEAALELLRQVYKPDGGIALHA SAEVDLDPSYRKFDPA" complement(9148..9849) /note="SAV5985" /codon_start=1 /transl_table=11 /product="putative TetR-family transcriptional regulator" /protein_id="BAC73697.1" /db_xref="GI:29609648" /translation="MTTGVRRMGVEERRQQLIGVALDLFSRSPDEVSIDETASAAG ISRLVTHYFPGKLSLFEALQORASDLQAGFVEFREGPLGARLLRVRRFPDFVEEH GPGFSALMRGGPAGVSSSTNALVDGVQAAAYVQILSHLVTEPPVRLVELVIRWSISLV ESTALIWLDRRIARGSEVQLVHDFAAALAAISAAYDDELAALLRDLVRDEAPDGPFS DLVRLIALTSPPTTA" complement(9928..11844) /note="SAV5986" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC73698.1" /db_xref="GI:29609649" /translation="MQSQDVAPRETTPASGPVVDVEQEAALVEHYPLRLVLAIVLPLP SLGNRRVLTAHALTQALPWRGRTSAVIPAQAKRGDPDGLYALVRLVRLTALEAGL PLGMAWPKRSQPLPLPGLRSLTFRSGGADLQDLQSLASPARAAHVLRGLE KLSDSVRRVLDIAGVADPAATAEAIAKVPAPDALDPSDFPCSLHARPTDLMRHQ HMAALAAAVAVCGALILPGGMAPDGAAPSAENPAARALDPAKLTAKTAATA WKTSAFTDFSWPARGRTGDTALLRELAIVARPGEAVRSATPGTSPGAPGPOL LYAGDVSAHVLLYDGLRIARYAEKDGTTGAALDFARVDGASGAREALVIGRAD NRYLTPAPWVGAERDLMKSPASGMOLNLTGVAAPLAGSAQRTGACTSNALQLTD GSGTRLTDLGELIPAHLTTPSPAPREASRTALHSMWPAFCSLASVRSQGVRSVNA WQQAQALPDGSGSAGVCTRADTWKGGTGVLAQFHTPGGAVGAVAAKAEDSPACGA RDPHLVAGLVKSGTGAWYLLAAGSSDTSIRATGGIDGAEGNLLVTRTERGARAGL KGVLDGRPTFSGLR" complement(12002..12514) /note="SAV5987" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC73699.1" /db_xref="GI:29609650" /translation="MGVPVTENWDERVAWAATPDDYAEKDAADFRAVVDALVAELPA DSPFPPEQACAMDSGSDKAVPLRYREALARGLSYGRRAKTIQLSSSLRNITQAAE GVKLLTPELGPSPDELDDAVRACLALCLADLGRERGLSLVLGALAPHLPRYQSRMAN YARLLVPEPD" complement(12571..15687) /note="SAV5988" /codon_start=1 /transl_table=11 /product="putative alpha-N-acetylglucosaminidase"	278568 GCACACGCATGGCGCTCGGATCTCGAGGCTCCGCGAGCGGTCTCGGCACCATGTC 278509 41 luTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleLeuGlnP 61 278508 CCCTCGAGGCAACACACCGACCGCTACGGG-----CTGCTCAGCGCGCGC-CTCCGCC 278456 61 roGlnArgArgProSerLeu-----LeuSerGluPheGlnProG 75 278455 GTCTCCGCCGAGACCTCGCTCGCTCCATGCTCCACCGCGGCGAGCTCCAAGCTC 278396 75 lyAsnGluArgSerGlnGluLeuHisLeuArgPro----- 86 278395 GCCGTGGCGCTCGACCTGAACTGCACGCACACCGGGCGTACGTCGGGCTCGTCAAC 278336 87 --GluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerL 106 278335 GGAGTGGCCACACCCGTGCACCGGGGGCGTCCACGGGAGCTACGATCGTATCAGC 278276 106 ysArg-----ProArgLeuGluLeuLeuProAsp---ProLeuLeuArg---ProS 121 278275 GACGAGGAGGCAAGCGGTGTGCACCGCGCTGACCTGCTGCTGCGGAACTCCCC 278216 121 erProLeuLeuAla-----ThrGlyGln----- 128 278215 GCGAATAGCGCGCTTCGGGGGCGGACCGGGCGGGAAGGACGGGACCGGCGG 278156 129 --ProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluProV 148 278155 TACCGCGGGAACCGGACCTTTGGCCCGACA--CGAAGATCAACCGC---CTCGGGCCCG 278100 148 alSerProSerProProHisThrAspProGluLeuGluLeuValProProArgLeuS 168 278099 TCGAACCGGACTTCCACCGCGCGCGCGCC-----CTCACCGCGCGCGCGACA 278049 168 erLysGluGluLeuIleGlnAsnMetAspArgVal---AspArgGluIleThrMetValG 187 278048 CGCGCTCGGGG-----CGAACTCCCGCGCTGCGCGGACCGTGGCGCCACACCGCCG 277995 187 luGlnGlnIleSerLysLysLysLysGlnGlnGlnGluGluGluAlaLysP 207 277994 TCCACACCGCGCGCGCTCCAGCGCGTCCCGCGCGCGCTG---GCGACGGCGCGCGCC 277938 207 roProGluProGluLysProValSerProProProIleGluSerLysHisArgSerLeuV 227 277937 CCGCGCGCGCGCGCGACCTTACCAGC-----TCTCGCACCGCGCACGACA 277893 227 alGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIle----- 244 277892 CCCTCCCGCGCACTGACAGCTCAAGCGCGCGCGCGCGCGCTCGCGCGCGCATCTCC 277833 245 -----LeuGluGluGlyProGln-ValGluLeuProLeuTyrAsnGlnProSer 261 277832 GTACCCGACTCGCGCGCGCTCCCGCGCGGTCAACATCCCG----- 277791 262 AspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLeuLeuIle 281 277790 -----CGCCTGGACCTGATCTCGCGCACCCCAAGCAGCAGCGCGCTACAGCTTC 277740 282 LeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCys----- 299 277739 CTGCTCGCGCGCAACCGCGCACCGCTCTCCCGGTGCGCGCACCGCGCTGCCCGCCA 277680 300 GlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsn 319 277679 CAGGAAAGTAACTACTGTAACTGTCATGCAATCAAGG-----AAT 277638 320 ProArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGlu 339 277637 GCGCGTCCGCGCGCTTGGCAACGCTC-----CCTCAGACCGCGCGC 277596 340 IleArgLysGlnArgGluLeuGlnGlnArgMetGlnSerArgValGlyGlnArgGlySer 359
-----	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Alignment Scores:

Pred. No.:	0.000253	Length:	299925
Score:	665.00	Matches:	640
Percent Similarity:	30.98%	Conservative:	255
Best Local Similarity:	22.15%	Mismatches:	1090
Query Match:	5.03%	Indels:	918
DB:	1	Gaps:	126

US-09-522-753-5 (1-2517) x AP005045 (1-299925)

Qy	2	SerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluPro-ArgTyrProPr	21
Db	278628	TCGACGAGTAGCGCGGCTCGCGCTCGACCTGCGCGCGCTCTCTCGCGCGGACACCTCG	278569
Qy	21	oHisSerLeuSerTyrPro-ValGlnIleAlaArgThrHisThrAspValGlyLeuLeuG	41

Db 277595 CCAGCCCCCGCCAGCCCTCGCAAGATCACTTGGAGCCC-----CTGAGCGGAACA 277545
Qy 360 GlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleLeuLeuLeuLeuLeu 379
Db 277544 GGGGTTCTCACCATGGGT-CACGGGGAGCAGCGGCCAGGCCCCCGACCGACCGCCC- 277487
Qy 380 SerGluGlnGlnLeuGluLeuGlnMetArgGlnLeuAlaValIleProProMetLeu 399
Db 277486 CCCACCCCGGACCAACCCAGCGCATACAGAGCGGTACAGCTGTGCCCATCTCCCTC 277427
Qy 400 TyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPro 419
Db 277426 CCCGACGCC----- 277418
Qy 420 MetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThr 439
Db 277417 -----GCCACCTGGCGTTATAGTCAACGCCAGTCAACCGCGC 277382
Qy 440 PheArgGluLysPheMetGlnHisPro-----LysAsnPhe 451
Db 277381 CGCGGGCGGCTGTGCAGACGCGGCCAGCAACCCAGTACGGCCCGCGGAGACACAG 277322
Qy 452 GlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyr 471
Db 277321 GCCCTCCAGCAGGAGGTCCGCCAGGAGGAGTGTGTCGCACACCGTTCTTGTCT 277262
Qy 472 LeuThrLysLysAsnGluAsnTyrLysSerLeu-----ValArgArgSerTyr 487
Db 277261 GATACTCACCACCGTGTCAACCGGAGCACTCACGCTCACCGCTTTCGGGATCGCGCGA 277202
Qy 488 ArgArgArgLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 507
Db 277201 CGACGACAGGGCAGCAGCAGCGGACGACACACCGTGTGTCATCGCGGTGACGCGGCC 277142
Qy 508 GlnGlnGlnPro-----MetProArgSerSerGlnGlnGluLysAspGluLys 523
Db 277141 CCTCAGCGCCAGAACTCCGCCACCGGCTCGGCATCCAGTACGGCTCCAGATCGCGCT 277082
Qy 524 GluLysGluLysGluAlaGluLysGluGluLysProGluValGluAsnAspLysGlu 543
Db 277081 CGACGA-----CGCAACAGAAAGAACTCGTGCCTCGCGCGCT- 277046
Qy 544 AspLeuLeuLysGluLysThrAspThrSerGlyGluAspAsn-AspGluLysGluAl 563
Db 277045 -----CACGTTCAAGTCAAGCCCTCGACGACAG--GC 277013
Qy 563 aValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleTh 583
Db 277012 CGTCCGCGCAGCGCCAGCAGACCGCCAGCGCTCGTCCCAACAGGACGCTCTC- 276955
Qy 583 rArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaG1 603
Db 276954 ---GGCGTCTGCGCGCCGCTGAACTCCGGTGTGCGCCACCCAGATGACGAGGCTTCGA 276899
Qy 603 uLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAl 623
Db 276898 CACGCCCGGCTCTCGATCTCCCGTCGACACACCGCCCGCGAG----- 276853
Qy 623 aLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySe 643
Db 276852 -----CTGACCCAGGGCAAGAACTGGCAGAGC----- 276826
Qy 643 rLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAs 663
Db 276825 -----AAGAGGAACG-TCCTTACA 276807
Qy 663 pGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgLysLys 683
Db 276806 AGACGTACTTCGCGACCGCCACCGAGCGCTCCAGGGCGGTTCGCGCGCGACTACG 276747
Qy 683 sLysLysAlaProAlaAlaAspGlnGluAlaAlaPheProProValValGluAspG1 703
Db 276746 CCTACAACACCT----- 276734

Qy 703 uGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAl 723
Db 276733 -----GAAGAAGCGCAAGG 276720
Qy 723 aLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAs 743
Db 276719 TCTTCATCG-TCGACGACAAGCAGACCT-----ACGGCGCGCGCTCCCAAG 276673
Qy 743 nAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThrG1 763
Db 276672 CTCCTTCAAGCAGAACTTCAAAA-----AGCGCGCGCGCAAG- 276636
Qy 763 yGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyProPr 783
Db 276635 -----TACCGCGCGAGGACACATCAGCGTGGCGACACGACTTCTCTCG 276589
Qy 783 oThrProProArgArgThrSerArgAlaPro----- 793
Db 276588 CTCGTCAACCAAGTCAAGAACTCCGGCGCGACCTTCTACTACGCGCGCAGTACGAC 276529
Qy 793 ----- 793
Db 276528 GAGTCCGAGAAGATCACCAAGCAGCTCAAGGACGGCGGCTCGACATCCCGTGTTCGGC 276469
Qy 794 -----IleGluProThrProAlaSerGluAlaThrGlyAlaPro----- 806
Db 276468 GGCAGCGGCATGTTCTCCGACACCTACATCAGACCGCGGCAAGCCCTCCGAAGCGCAG 276409
Qy 807 ---ThrProProAlaProProSerProSerAlaProProPro-----ValVa 822
Db 276408 CTGCTCACCTCGCGCTCCCGTCAACCCCTCCCGCGCGCCCAAGAGTTTCATCGCC 276349
Qy 822 lProLysGluGluLysGluGluThrAlaAlaAlaProProValGluGluGluG1 842
Db 276348 ACCTCAAGCGGAGAGTATACCCCGCGACTACGGCACTTACGGTGGTACTCTCTACGAC 276289
Qy 842 uGlnLysProProAlaAlaGluLeuAlaValAspThrGlyLysAlaGluGluProVa 862
Db 276288 GCGCCACCGCCATCATCAAGCCATCGCGCAACGCTCGTGAAGACGCAAGTTCCTCT- 276231
Qy 862 lLysSerGluCysThrGluGluAlaGluGluGluProAlaLysGlyLysAspAlaGluAl 882
Db 276230 -----CGACGCGCGCA----- 276219
Qy 882 aAlaGluAlaThrAlaGluLysAlaGluLysAlaGluLysLysGluGlySerGlyAr 902
Db 276218 -----CCAAGATCTCGACGAGGTCAG 276196
Qy 902 gAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSe 922
Db 276195 AAGACCAAGTTTCAGCGCATCGCGCGCGCTTCGACGAGTACGCGGACA----- 276141
Qy 922 rAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgPr 942
Db 276140 -----CCACCAACAGCAGCTCACCGTCTACAGGTCTCGAAGGGCAAGTGAAGGCC 276088
Qy 942 oSerLeuLeuThrProThr-----GlyAspProArgAlaAsnAlaSerProGlnLysPr 960
Db 276087 GTCAAGAGCGGCACCTTACACGCGCGGTGACCCCGCGCACCCACACGACCCCGGCGC 276028
Qy 960 oLeuAspLeuLysGlnLeuLysGlnArgAlaAlaIleProProIleGlnValThrLy 980
Db 276027 G-----CGCGG 276022
Qy 980 sValHisGluProProArgGluAspAlaAlaPro-ThrLysProAlaProAlaProp 1000
Db 276021 CACCCACACAGCGCCAGCGCGCGCGCTCCACTCAGTCTGCTCACTTCCACACCCA 275962
Qy 1000 ro---ProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerP 1019
Db 275961 CATGAGCGCGCGGTGCACACCTTCGCGCAGCAGCTGGCCCAACGCGGTGT-TCCTCGG 275903


```
Qy 1019 roArg-----GlyLysSerArgS 1025
Db 275902 CTCGATACGGGCTGATCGCATCGGTTACACGATGGTGTACGGCATCGTCCAGCTCAT 275843
Qy 1025 erProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyA 1045
Db 275842 CACTTCGCCACCGCGGAGATCTTCATGACCGGAGGCTTCGGCGCATCTCAGGFTACTT 275783
Qy 1045 spProProCysTriPThrSerGlyLeuProPheProValProProArgGluValLysA 1065
Db 275782 CTACGCTCTGCCAGCGGCACAT---CCATGTGATAGCCCTCCCGGATGCTCATCGG 275726
Qy 1065 laSerProHisAlaPro-----AspProSerAlaPheSerTyrAlaProProGlyHisP 1083
Db 275725 AGCGGCGCTCATCGCCGCTACTCATCGCGCTCGGGCGGAACGCTTCGCCCTACCGGC--- 275670
Qy 1083 roLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProProThrIles 1103
Db 275669 -----CCCTGGCGGCGCACACACGCCCTCGCCCCC-----T 275639
Qy 1103 erAenProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleG 1123
Db 275638 CATCACCGCCATCGGCTCTCCCTCGGCTCC----- 275607
Qy 1123 lyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysA 1143
Db 275606 -----AGCAGGTGGTCTTCACTTCTACCCGCGACGCCA-----A 275573
Qy 1143 laProValGlyProValThrMetGlyLeuProLeuPro----- 1155
Db 275572 GACCGCAGCAGACCTTCCCCAGCTGGGCACCGAGTCTCTCAACATCGGCTCCGFTCCACAT 275513
Qy 1156 -----MetAspProLysLysLeuAlaProPheSerGlyV 1167
Db 275512 CCAGAGCGGCAGCTCTTCATCATCATCGCGCGCCCATCGAGGCCACCGCCAGGACCGGACAC 275393
Qy 1167 alLysGlnGluInLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValP 1187
Db 275452 CTTGCTCGGCTCTCCCGCACCGCGCGCGCATCGAGGCCACCGCCAGGACCGGACAC 275393
Qy 1187 roThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlyS 1207
Db 275392 CGCCAGCTCATGGGCATCGACA-----CCAACCGCATCATCGTATCGCCCTTCGCCAT 275339
Qy 1207 erIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgG 1227
Db 275338 CGCGGACTCTTCGCCGCGTGGCG-----CGTGCCTACGCGCTCGCTACGSCCA 275285
Qy 1227 lySerIleThr-HisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIle 1246
Db 275284 GATCAAGTACGACATGGGCTTCCA----- 275261
Qy 1247 IleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGly 1266
Db 275260 -----GATGGGCTCAAGGCTTTCACCGGCGGT-----CCTCGGCGG 275222
Qy 1267 HisValIleTyrGluGlyLysGlyHisValLeuSerTyrGluGlyGlyMetSerVal 1286
Db 275221 CATCG-----GCAACATCTACG 275205
Qy 1287 ThrGlnCysSerLysGluAspGlyArgSerSer-----SerGlyProProHisGlu 1303
Db 275204 CGCCATGCTCGGGAGTCTCTCGCGGTGGCGAAACCTTCGCCACCGCTACATCG 275145
Qy 1304 ThrAlaalaProLysArgThrTyrAspMetMetGluGlyArg----- 1317
Db 275144 ACGGCATCCCGCGCATCAGCAGCTCGCGCGCGCAGCTGGGCGCCCTCGGCTCGGCTTCG 275085
Qy 1318 -----ValGlyArgAlaIleSerSerAlaSerIle----- 1327
Db 275084 TCCTCTCATCTCGTACTGCTGCTGCGGCGCACAAAGGCTCTCTCGGCGAACCGTCCGCG 275025
Qy 1328 -----GluGlyLeuMet 1331
```

```
Db 275024 ACAGGGCGTGTATCACCATGACCCCGCCGACCCAGCACCGAGAACCGCGCGGCGCT 274965
Qy 1332 GlyArgAlaIleProProGluArgHisSerProHis---HisLeuLysGluGlnHisHis 1350
Db 274964 CATCCCGCTCCCCAGCGCCGCGCCCTCTCTCATCGTATCGGCGCCCTCGGCACCAT 274905
Qy 1351 IleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyr 1370
Db 274904 CGCCAG-----CGCTTCATGTCTGGACCTGGACCGCGCG 274869
Qy 1371 LeuArgArgGluAlaLysLeuLeuLysArgGlyThrProProProPro----- 1387
Db 274868 CTTCCCGGCTGACCTCACCGCTTACGGCTACCCCGCGGCTTCAGACCCCTCGGCATCGT 274809
Qy 1388 -----ProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeu 1402
Db 274808 CGGGCGGTCTATCACCCTCTCTCGGCTCACCGGTGGGACATCCCGCGCTCGGCTG 274749
Qy 1403 Gly-----ProLeuLysLeuLysProAlaHisGlu 1412
Db 274748 GGCAACCCCGGAACGCGGCTCCCGCATCGCCTCGCGCGGCTCCGCTTCCACCGT 274689
Qy 1413 GlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGlu 1432
Db 274688 CTGCTGTGTACACCGCATCGCATCGCGTCTCGA-----CCTCGCGGCGCT 274644
Qy 1433 LeuArgHisThrProGluLeuProLeuAla-----ProArgProLeuLysGluGlySer 1450
Db 274643 CGCAACCTCGACCCCGGCG-CCTAGTCGCCATGGCGGCTCCCTCATCGCCTCCTCG 274585
Qy 1451 IleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLys 1470
Db 274584 GCGCCCTCGGCTCCCGACCCCGCGGCGCACCTGGCGG---ACTACCTCGCCAAAGCGG 274528
Qy 1471 HisAspValArgSerLeuIleGlySerPro-----GlyArgThrPhe 1484
Db 274527 AGAACATCCCGCGCGCGGCGGAACCTCGCCCTCGGTCAGCGCTCGTATCACCTCG 274468
Qy 1485 ProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyr 1504
Db 274467 CCACCTCTCGGCGCTGTATCGTCTTACCTACCTACGCGATCGGCG----- 274426
Qy 1505 GluGluSerLeuLysSerArgProGlyThrAlaSerSerGlyGlySerIleAlaArg 1524
Db 274425 TGAACGACGAGAGAGGAAACCTTCATCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 274366
Qy 1525 Gly---AlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyr 1543
Db 274365 GGGCACTCTTCGCGCGCGGCTCTCGACCGGTTCTCCAGATCAACGCGCAGCACA--- 274309
Qy 1544 GluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThr----- 1561
Db 274308 AAGGATTCGCCACACACCTGCTCTCTCGCGCGGATCTTCCCTTCACCCAGAGCG 274249
Qy 1562 -----MetArgGluProThrProArgLeu 1569
Db 274248 ACGACCAACGCAACCTCGGCTGACATCTGTATCGTCCGGAACGCTCGGCC----- 274195
Qy 1570 GlnGluGlySerLeuSerSerLysAlaSerGlnAspArgLysLeuThrSerThrPro 1589
Db 274194 -----TCGGCTCAACATCGTCTCGGCTTCACCGGACTCTCGACCTCGGATACGTCG 274141
Qy 1590 ArgGluIleAlaLysSerProHisSerThrValProGluHisProHisProIleSer 1609
Db 274140 CTTCTCTCGCGCTCGGCGCT-----ACGCGCGCGCTCTCTCTCGGCTCGAGTTCT 274087
Qy 1610 ProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIlePro 1629
Db 274086 CCGCTTCGCGGCTCCAGTTCCCTCTCTGGG-----CAGCGGCACTGACCG 274039
Qy 1630 LeuAlaPheAspPro-ThrSerIle-----ProArgGlyIleProLeuAspAlaAla 1647
```

Db 274038 ---GCATGGGAGCCTCACTGCTTCGCGCTCTCATCGCGCCCC---GACGCTCCGGC 273985
Qy 1647 aAlaTyTyLeuProArgHisLeuAlaProAn----- 1658
Db 273984 TCGCGGCGACTACCTCGCCGCACTGCTCAGCTCGCTCGGTGAGATCTTCGCGATCGCGC 273925
Qy 1659 -----ProThyTyProHisLeuTyProProTyTyLeuLeAr 1671
Db 273924 TCAACAGCCTCGAGGCTCTCCGGACCCAACTCACCACGCGCCCAA-----CG 273874
Qy 1671 gGlyTyProAspThrAlaAla-----LeuLuAsnAr 1682
Db 273873 GCATCTCCGGATCCCGACCTCGAAATTCGATTCGAATTCACCTCGGGATCTCCACGAGC 273814
Qy 1682 gGlnThrIle-----IleAsnAspTyTyIleThrSerGlnGlnMetHisHisAs 1698
Db 273813 TCGGCTCTTCACCGTGGCGGCTTCGGAATCTCTCTGATGATCATCAGCG 273754
Qy 1698 nThrAlaThrAla-MetAlaGlnArgAlaAspMetLeuArg-----GlyLeuS 1714
Db 273753 CACTCGTGTGCTGCTCTTCCACGAGGCTCGGACTCCCGACTCCCGCTCGCGCTCTCGGATCG 273694
Qy 1714 erProArgGluSerSerLeuAlaLeuAsnTyAlaAlaGlyProArgGlyIle---IleA 1733
Db 273693 CCATCCGCGAGGACGAGCCGCCACC---GCCATGGGATCAACGGCTTCGCGGTCA 273637
Qy 1733 spLeu-SerGlnValProHisLeuProValLeuValProProThrProGlyThrPro--- 1751
Db 273636 AACCTATCGCTTCGCGCTCG-----CGGCTCTCTCGCGGCTCGCGGTGTA 273589
Qy 1752 -----AlaThrAlaMetAspArgLeu-----AlaTyR 1760
Db 273588 CGGTGAGCGCCACGCTCGGCTACAGCTCAACCGCGCGGTACCACTTCGCGGCTCGC 273529
Qy 1761 LeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGly 1780
Db 273528 TGCCCGCGAATCCGCTCTCTGCTCGCGGCTGCTCTCGCGGATCAACGGCTCGCCTCA 273469
Qy 1781 GlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgAspArg 1800
Db 273468 ACGGCGCCATCTCGCGCCACCTCTCTACCTCATCC-----CCGAGA 273424
Qy 1801 AspArgGluArgAspArgAspArgGluArgGlyLysSerIleLeuThrSerThrThr 1820
Db 273423 AGCTCTCTCTCTGAGGAGTACAGCTCTCGCTCGGATCGCGCTCTCTCTCTCA 273364
Qy 1821 ValGluHisAlaProIleTrpArgProGlyThrGluGlnSerSerGlySerGlySer 1840
Db 273363 TCGGCTTCGCGCCCG-----AAGGCATCATCGCCA 273334
Qy 1841 SerGlyGlyGlyGlySerSerSerArgProAla----- 1852
Db 273333 ACCGCGCGCCAGCTCGAAATTCACGAGACCGACCATGACGTACCGGAACACGCC 273274
Qy 1853 -----SerHisSerHis 1856
Db 273273 TGACCGAACCCCGCTCGGCTCAAGAGCGGGCGGTGACCGACATGACACCAACAC 273214
Qy 1857 AlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSer 1876
Db 273213 GGCCACCTCCACCGTCTCGAAGCTTCGCGCTCGCCTCGCCTCGCGG----- 273163
Qy 1877 valLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThr 1896
Db 273162 ---ACTCACCGCGCTACGCTCCGT-----CGACCTCACCGCTCA-ACA 273125
Qy 1897 ValLeuArgSerThrSerSerSerProValArgProAla-----AlaThr 1912
Db 273124 CGCGCGAATCGTCCGCTCATCGGCCCAACCGCGCGCGGACCACTCTTCTCAACT 273065
Qy 1913 PheProProAlaThrHisCysPro-LeuGly---GlyThrLeuAspGlyValTy-Proth 1931
Db 273064 GCCTCACCGGCTCTACGTCCTCCCGACCGAGGGCAAGGTCTCTTACAGGGCAAGTCTCTG 273005

Qy 1931 rLeuMetGluProValLeuLeuPro-----LysGl 1941
Db 273004 CGCGAAGCCCCA-----CCTGTGACAAGCGCGGCATCGCCGCGACGTTCAG 272954
Qy 1941 uAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeuAl 1961
Db 272953 ACATCCGGCTCTTCGCCCAACATGACCGCTCTCGA----- 272920
Qy 1961 aLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluPr 1981
Db 272919 -AAACGCTCTCGTAGAGCGCACACAGGACCAAGAGAGGCTCTGTCGCGCTCTCTG 272861
Qy 1981 oArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAs 2001
Db 272860 GCGGCGCGGCTTCG-----CAAGCGCAAAACGCTCCCGGAACGGCCATGAAC 272807
Qy 2001 nLeuAlaProHisAlaSerProAsp-----ProProAlaPro----- 2014
Db 272806 TGCTGAGTTTCATCGGCTCGCCCAAGCGCGACCTCTCCCGCAACCTCCCTTACG 272747
Qy 2015 ---ProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIl 2033
Db 272746 GCGAACAGCGAAGCTCGAGATCGCCCGCGCACTCGCAGCGAACCGGCTCTGTCTCC 272687
Qy 2033 eGlnGluLeuGluLeuArgSerLeuGlyTyHisGlySerSerTySerProGluGlyVa 2053
Db 272686 TCGA-----CGAGCGCGCGCGCGCAT----- 272665
Qy 2053 lGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeu----- 2070
Db 272664 -GAACCGCGAGAGACGCGAGC-----GACCGAGGACTGTGCTTCG 272624
Qy 2071 ---ProLysHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLy 2089
Db 272623 CATTCGCGCAT-----GGCATCGCGCTCTCTCG 272594
Qy 2089 sGlnProGlyProValLysLeuGlyGlyGluAlaHisLeuProHisLeuArgProle 2109
Db 272593 TCAT-----CGACGACGACATCGCTTCTCATCTTCAACCTCT 272558
Qy 2109 u---ProGluSerGlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGl 2128
Db 272557 GCACCGCTCGCGCTCT-----CGTCCAGGCGGAGAGC 272522
Qy 2128 yHisGlnArgValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyTh 2148
Db 272521 TCATCGAAG-----CACCC-TGGAGGTCTGTACAGCGCGACGA-GC 272483
Qy 2148 rArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTy-Ser-PheProGl 2168
Db 272482 CGCTCATCG-CCGCTTACCTCGCGCAACCTTTCGAAGGCGCC----- 272441
Qy 2168 yAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyLeuPro----- 2186
Db 272440 -CGCGCACCGCGCGCGCGAGTCTGAAGCGCGCAAGCCCGCAAGCGCGGAG 272382
Qy 2187 -----ProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGl 2203
Db 272381 GCCGAAGCGCGCGCGCGCGCGAGCGCGCGAGCCAGCCACCGCGCGCGCGAGC 272322
Qy 2203 yLysArgSerProGlu-----ProAsnTyThrSerValLeuGlyGlyGlyGluAspGl 2221
Db 272321 ACCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 272270
Qy 2221 yIleGluProValSerProProGlyMetThrGluProGlyHisSerArgSerAlaVa 2241
Db 272269 -----CCTCAGGCTCTCTTACGCGCAAGATCGAAGCGCTCAAGG-- 272231
Qy 2241 lTyProLeuLeuTyArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSe 2261
Db 272230 -CATCTCTCTCAGCTCGAAGCGCGGCC-AGGTCTGTACCTCTCA-----TCGGCAC 272182

Qy 2261 rProGlyAsnThrSerGlnProProAlaPheSerLysLeuThrGluSerAsnSerAl 2281
 Db 272181 AACGGCGCTGGCAAGACGACACCTCTCCGGCTCTCAAGCCGACCGCG 272122
 Qy 2281 aMetValLysSerLysGlnGluIleAsnLysLeuAsnThrHisAsnArgAsnG1 2301
 Db 272121 GCGCGCATCGTCT-----TCGACGGCCACACGCTCA-- 272091
 Qy 2301 uProGluTyAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrG1 2321
 Db 272090 -----CCAGCG----- 272085
 Qy 2321 yThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetG1 2341
 Db 272085 ----- 272085
 Qy 2341 yLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerPr 2361
 Db 272084 -----TCCCC 272080
 Qy 2361 oProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaIle 2381
 Db 272079 GCCC-----ACAAGATCGTCGCGCTGGGCTG 272053
 Qy 2381 tProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyG1 2401
 Db 272052 GCCCACTCCCGAGGCGCGCACATCTCCCGCGCTGACCATCGCGGAGAACCTCCAG 271993
 Qy 2401 yLysAlaLysValSerGlyArg-----ProSerSerArgLys----- 2413
 Db 271992 CTGGAGCGTTCCTCCCAAGGACCAAGCCCGCATCGAAGGACATCCAGCGCGCTAC 271933
 Qy 2414 -AlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerVa 2433
 Db 271932 GACCTCTTCCCATCTCTCGGGGAAGCA--GGNAGCAGCGCCGCGAACCTCTCGGGC 271876
 Qy 2433 lHisSerGluGly-----AspCy 2439
 Db 271875 GCGCAACAGCAGATGCTGGCCATGGCGCGGCTGATGTCACGCCAACTGCTCATG 271816
 Qy 2439 sAsnArgArgThrProLeuThrAsnArgValTrpGluAsArgPro----- 2454
 Db 271815 CTCGACGAGCCCTCCA-----TGGGACTCTCGCCCATCATGATGCAAG 271771
 Qy 2454 ----- 2454
 Db 271770 ATCATGCCACCATCGCGGAGCTCAAGTCCAGGGCAGCAGATCTCTCTCGTGAACAG 271711
 Qy 2455 -----SerSerAlaG1 2458
 Db 271710 AACGCCCGAGCGGCGCTCTCACTCGCGCACCGAGGCCACGTCATGGAGATCGGCAACATC 271651
 Qy 2458 ySerThrProPhePro-TyrAsnProLeuIleMetArgLeuGlnAla-----G 2474
 Db 271650 GTCTCTCCGGACCGGACGAGGACCTCTGCACGACAGTCCGTGCGCAAGGCGTACCTG 271591
 Qy 2474 yValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyP 2494
 Db 271590 GCGGAGGACTGAGCCCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 271531
 Qy 2494 roHis 2495
 Db 271530 CGCAC 271526
 RESULT 90
 BX842580
 ID BX842580 standard; circular genomic DNA; PRO; 346051 BP.
 XX
 AC BX842580; AL008687; AL021185; AL021186; AL021246; AL123456;
 AC 277250; 277724; 280225; 283863; 295387; 295556; 296072; 297051; 298045;
 AC 298209;
 XX
 SV BX842580.1

XX
 DT 21-NOV-2003 (Rel. 77, Created)
 DT 21-NOV-2003 (Rel. 77, Last updated, Version 1)
 XX
 DE Mycobacterium tuberculosis H37Rv complete genome; segment 9/13
 XX
 KW complete genome.
 XX
 OS Mycobacterium tuberculosis H37Rv
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;
 OC Mycobacterium tuberculosis complex; Mycobacterium tuberculosis.
 XX
 [1]
 RN MEDLINE; 98295987.
 RX PUBMED; 9634230.
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry III C.E., Tekala F., Badcock K.,
 RA Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K.,
 RA Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K.,
 RA Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J.,
 RA Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S.,
 RA Squares S., Squires R., Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the complete
 RL genome sequence";
 RL Nature 393:537-544(1998).
 XX
 [2]
 RN PUBMED; 12368430.
 RA Camus J.C., Pryor M.J., Medigue C., Cole S.T.;
 RT "Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv";
 RL Microbiology 148:2967-2973(2002).
 XX
 [3]
 RN 1-346051
 RP Parkhill J.;
 RA
 RT Submitted (11-JUN-1998) to the EMBL/GenBank/DBJ databases.
 RL Submitted on behalf of the Mycobacterium tuberculosis sequencing and
 RL mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton,
 RL Cambridge CB10 1SA Unite de Genetique Molculaire Bacterienne, Institut
 RL Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail:
 RL parkhill@sanger.ac.uk
 XX
 CC Notes:
 CC Details of M. tuberculosis sequencing at the Sanger Centre
 CC are available on the World Wide Web.
 CC (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/)
 CC
 FH Key Location/Qualifiers
 FT source 1. 346051
 FT /db_xref="taxon:83332"
 FT /mol_type="genomic DNA"
 FT /organism="Mycobacterium tuberculosis H37Rv"
 FT /strain="H37Rv"
 FT 142. 528
 FT CDS
 FT /note="Rv2470, (MTV008.26), len: 128 aa. Possible glbO,
 FT globin-like protein, highly similar to Q9CC59[GLBO]ML1253
 FT HEMOGLOBIN-LIKE (OXYGEN CARRIER) from Mycobacterium leprae
 FT (128 aa), FASTA scores: opt: 767, E(): 4e-47, (88.1%
 FT identity in 126 aa overlap); Q9X7B3[MLCB1610.14c PUTATIVE
 FT GLOBIN from Mycobacterium leprae (131 aa); Q9L250[SCGD10.14
 FT PUTATIVE GLOBIN from Streptomyces coelicolor (137 aa),
 FT FASTA scores: opt: 466, E(): 5.7e-26, (53.8% identity in
 FT 125 aa overlap). Also similar to O31607 YJBI PROTEIN from
 FT Bacillus subtilis (132 aa), FASTA scores: opt: 294, E():
 FT 6.6e-14; (39.85% identity in 128 aa overlap). COULD BELONG
 FT TO PROTOZOAN/CYANOBACTERIAL GLOBIN FAMILY PROTEIN."
 FT /transl_table=11
 FT /function="OXYGEN CARRIER, INVOLVED IN OXYGEN TRANSPORT."
 FT /gene="glbO"
 FT /locus_tag="Rv2470"


```
Qy 85 -----ArgProGluSerHisSerTyrLeuProGluLeuGlyLysSer 98
Db 25936 GCAGGGCCCGCGGCACGCCACACGGCTGCAGACATGATGCCAACAGCAACGCGTGGC 25995
Qy 99 GluMetGluPheIleGluSer-----LysArgPro-----ArgLeuGluLeu 112
Db 25996 CGGTATCGCCATTCTCGCGGTGTAACCGAAGCGCCCGGTAGGTTCGTGATTCGGA--- 26052
Qy 113 LeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySer 132
Db 26053 -----GCTCGCCCTGGTTGAGCCCGCGTTCATGCGCGCCACCGAGGGGG 26097
Qy 133 GluAspLeuThrLysAspArgSerLeuThrGly-----Lys 144
Db 26098 CGTCAAGCCCGCGGCAACCGCCGCTAGTGGTCTCGCGTGTGCGGCACGCGCATCGG 26157
Qy 145 LeuGluProValSerProProSerProProHisThrAsp---ProGluLeuGluLeuVal 163
Db 26158 CATCGCGGATTCACGCGAGCTTCTCCAAACGATACCTGCGCGCACCGCTCTAGCAGCGGT 26217
Qy 164 ProPro-----ArgLeuSerLysGluLeuIleGlnAsnMetAspArg 178
Db 26218 AGCGCGTTCGCGCTGCTGTCGCTGCGTCAACACAGAGACTTGTCTGCCAGCAGGCTGA 26277
Qy 179 ValAspArgGluIleThrMetValGluGlnGlnIleSerLysLysLysLysGlnGln 198
Db 26278 GCAGATCGACCACTCTGAGCGCTGAACGTACCGCGCGCGGTGCGCTTGGGCACCGT 26337
Qy 199 GlnLeuGluGluAlaLysProProGluProGluLysProValSerProProPro 218
Db 26338 CGAGATCAAAACCCGCTCGGGAACACCGCATCGCCGAA-----ACAGCACTGTCTCG 26391
Qy 219 IleGluSerLysHisArgSerLeuValGlnIleIle-TyrAspLysAsnArgLysLysAl 238
Db 26392 GTC-----CGGTACGACGCGCATGTGACCATGTGACAGCAAGACCGCATCGTCT 26439
Qy 238 aGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAs 258
Db 26440 GCTGCGCGGCACGCAATACCGCATCCCGGTACAGCGCGGAA---CGGTATCGCA 26496
Qy 258 nGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLy 278
Db 26497 AGCTGTGACGATTTTCGTGTCAG----- 26518
Qy 278 sLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPh 298
Db 26519 -----CGCCAGGCGACGCAACCCGCGACGC----- 26542
Qy 298 eCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGlu----- 314
Db 26543 -TGC---AAGTTCGATCGCAGCGGATGCGTGCAGTCCGTGGCAGATCTCGGTACCA 26598
Qy 315 -----ArgIleGluAsnAsnProArg----- 321
Db 26599 GGGCGAGGTTGTCGCGCAGTATCTCGAGTTTCGGCGCGCTCACAGCGCGGTGCGTGA 26658
Qy 322 -----ArgArgAlaLysGluSerLysValArgGluTyrTy 333
Db 26659 ACAACTCGATCGCTCGCCGTGCGCCAGCGGGGAACCCGCCAAATCTGTCTACCCGCCA 26718
Qy 333 rGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerAr 353
Db 26719 CCGCGATCGGTTCCCGCTGCTGCGCAATACCTCAGCGGTGGGCAGCGCCCGGACGACG 26778
Qy 353 gValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHis---GluVa 372
Db 26779 CGAGCATCAGGCGCGCACCGGTGCGAGCAAGTCTCGCAGTGTCTCCAGCACTACCAGCA 26838
Qy 372 lSerGluIleIleAspGlyLeuSerGluGlnGlnAsnLeuGluLysGlnMetArgGlnLe 392
Db 26839 TGGCGCGGTGCGCGAT-----ACGCCGCAAAATGGGTGTCCACGTCGACGCGCGCGGT 26892
```

```
Qy 392 uAlaValIlePro-----ProMetLeuTyrAspAlaAspGlnGlnArgIle----- 407
Db 26893 GATCCGGCAACCCCAAAACCCGCGCGCGATCGGCACACAGCGCGCGGTTCGGTGATCG 26952
Qy 408 -----LysPheIle 410
Db 26953 GCGCCAGTTGACATACAAACCCCGTCCGATTAACCGTCCGGCAACGCGCGTCCGACCT 27012
Qy 410 eAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAs 430
Db 27013 GTGTCCGAGCGGTGCTTTCAGACCCCGCGGACACCGGTAAAGGTGACCCACCGTTGA 27072
Qy 430 nMetTrpSerGluGlnGlnLysGluThrPheArgGluLysPheMetGlnHisProLysAs 450
Db 27073 CGTC-----CAGCAGCCACCGGACTTTCGCGCACTTCTGTCAGCGCGCCACCA 27120
Qy 450 nPheGlyLeuIleAlaSerPheLeuGluArgLysThrVal----- 463
Db 27121 G-----CCGAGTGAGTGGCGCGAAGACAGTAGTCCGCCAACGACTTTCGGG 27168
Qy 464 -----AlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyrLy 480
Db 27169 TCCGACGCGCGGAAACCGCTTGTGCAGATCAGGTGACAGCTGCACCCCGTTCG 27228
Qy 480 sSer-----LeuValArgArgSerTyrArgArgGlyLysSerGl 494
Db 27229 GTCCGGGCGAGTCTGTCAGCGCGGTAGGTACCGAGGTGCTTTCAGCCACGCTCTTGGCA 27288
Qy 494 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 514
Db 27289 GCAGGTGAGCAACACAGATCTGCTGTAGTTCCTCCAGCAACACAGCTCTGCGCCCGG 27348
Qy 514 gSerSerGln----- 517
Db 27349 GTCGCGCACCGCGCGGTGCGGTGCTGATGTGCGGCTTCCAGCTGCTCCCTCTCGCGGTG 27408
Qy 518 -GluGluLysAspGluLysGluLysGluLysGluAlaGluLysGluGluLysPro-- 536
Db 27409 ACGACACCTCCCGCGGTGTCATCCGATGCGCAGCGGATCGGTGCGCGCGCGCGCT 27468
Qy 537 -GluValGluAsnAspLysGluAspLeuLeuLysGluLysThrAspThrSerGlyGl 556
Db 27469 GCAAGCCCGAGCGCGCACCGCACCGCGCTC-----GGATCGCGCGCGCA 27510
Qy 556 uAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThr-----Al 572
Db 27511 ACCCCACCAAGAGTGTGCTGCTTCCGCTTGTGCGACCGCGGCAACCCCGCGGTGCTCGC 27570
Qy 572 aAsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSe 592
Db 27571 GAACCAATTCCGTGAGGTTTC-GGTCCAGTTTCGCGATCCCGTTCGTCGTCGTCGAGAC 27629
Qy 592 r-----GluGluAlaIleThrProGlnGlnSerAl 602
Db 27630 CCGCGCAGGTGGTTCGCGCCTCGATATCGCGCCAGCAGCAACGTCACCGTCCGCTCGGT 27689
Qy 602 aGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGlu----- 618
Db 27690 ACAAGTCTGCTCACCATCTGCGTCCAGTCCACAGGT-ACCACGTGACGCGCGGGGTG 27748
Qy 619 -----GluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyAr 632
Db 27749 AATCTTGTCTATGACGACATCGAGCCAGCGGTAGCGCATATACATCGGCAC----- 27803
Qy 632 gAsnTrpSerAlaIleAlaArgMetValGlySerLysThrVal-----SerGlnCy 649
Db 27804 -----CTGCGCTAGATTGCTCGAAATCTTTGGCGCGCGGTCCATGTG 27847
Qy 649 sLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGlu----- 664
Db 27848 TTCTACGCGCTTTAGTTCGATGTCATTTCGCGACCGCGGTGCTGCTGCGCGCGGACCTACAGT 27907
Qy 665 ----IleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaAlaArgLysLy 683
```

Db	27908	GCCTGCTCTCGCTGGCAATGTCAGTCCCGCAGTGCCTGGCAGCATCCCGCCGCAAGAA	27967	Db	28876	CGTTTCCCGCGCGCGCCACCGCCGTCAGCCAGTCCGCGCGCGCTGCGCGCGCTGC	28935
Qy	683	sLys-----LysAlaProAlaAlaAlaSerGluGluA	694	Qy	1005	-----LeuGlnProGluSerAspAla	1011
Db	27968	CCGACACGCGCATCGTGGCGGTGCGTCGAAGTCGCGGCTACCGATCGCGGAGT	28027	Db	28936	CGCCCGCTGTTCTGTAATGTCAAAGCCGTCAACGCCGTTACCGCGGACCCACCAGCC	28995
Qy	694	la-----AlaPheProProValValGluAAspGluGluMetGluAlaSerGlyV	710	Qy	1012	ProGlnGlnProGlySerSerProArgGlyLysSerArg-----SerProAlaPro	1028
Db	28028	CACCGCGCGATGGGTTCGGAAGCTAGGGAATGATGACGATGGGCGCGCTCGG	28087	Db	28996	CGCCTTAGCTCTCGCGCCCGCGCTCACACCGTCCCGGTCTGAGTTCGCGCGCGCCA	29055
Qy	710	al-----SerGlyAsnGluGluMetValGluGluAlaG	722	Qy	1029	ProAla-----AspLysGluAlaPhe	1035
Db	28088	CGCCTTCGCGTAACCCCAACCATCGGAAACGAGCCTAGCTCGCGCGCGCGCA	28147	Db	29056	CGCGCCCGCGCTGCGCTCCCGCCACCGCTCCGCGCTCGAAGCCGCTTT	29115
Qy	722	luAlaLeuHisAlaSerGly-----AsnGluValProArg-----GlyG	735	Qy	1036	AlaAlaGluAlaGlnLysLeuProGlyAspPro-----ProCysTyrThrSer	1051
Db	28148	GAGCGACCATCGCGTGGCGCCCAACACGACGAGGCGATCCGGAATCTCTGACCATGGT	28207	Db	29116	AGTCTCGAGTCTGGGCGCAACCCCGCGCACCGCGCGCGCTGCTCTCTGGCGG---	29172
Qy	735	luCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProH	755	Qy	1052	GlyLeuPhePheProValProProArgGluValIleLysAlaSerProHisAlaProAsp	1071
Db	28208	GGTGTCTGCTACGTACGTTA-----ACGGAGATGGAGGCGCGCTTCGCGG-	28257	Db	29173	-----CCGCTTCGCGCGCGCTTCATCCCGCGCTCGCGCGCGCGCGCG	29220
Qy	755	isThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyA	775	Qy	1072	ProSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThr	1091
Db	28258	-----CCTTCACCGCGG-----	28269	Db	29221	-----TGTCCACCAACCCCGCGCGCGCGCGCGCGGCTG-----	29253
Qy	775	laAspGlyProProProGlyProProThrProProArgThrSerArgAlaProIleG	795	Qy	1092	AlaArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSe	1111
Db	28270	--GAACCGCGGAGCCAGGTGCGCTCCCTGTCGCGGAGCACCGACTCCCGCGAGC	28327	Db	29254	-----CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC	29295
Qy	795	luProThrProAlaSerGluAlaThrGlyAlaProThrPro-----ProProAlaProp	813	Qy	1111	rAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerVa	1131
Db	28328	AGCGAGCGCGCGGAGTCAAGCCACCGCACCGCGTCCGCGTCACTCCGCGCGCG	28387	Db	29296	CGCGCTGCGCCAGAAATGCTCGCGCGCGC-----	29326
Qy	813	roSerProSerAlaProProProValValProLysGluGluLysGluGluThrAlaA	833	Qy	1131	lGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGl	1151
Db	28388	CGTCCCGCGCTCACCGCGCGATGTGCG-AGGCGGAGGGCGCGGAAGCGCGCGAG	28446	Db	29327	-----CACCAGCCCCACCGCGCGCACCA-GCCCAACCGTTCGCCCGCAACCGTGAG	29378
Qy	833	laAlaProPro-----ValGluGluGlyGluGluGlnLysProProAlaAlaG	849	Qy	1151	yLeuPro---LeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGl	1170
Db	28447	CGCGCGTCCCGCGTCCGCTCACACCGCGGCGCGCACCGCTCCGCGCGCGCCCA	28506	Db	29379	CCCTCGAGGGCACCGTGGCGCGCTCGCCACCTGTCGCGCGCTCA-----	29424
Qy	849	luGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluG	869	Qy	1170	uGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGl	1190
Db	28507	CCCAAGCGCGGTI---GCCGCGGTGCCACCGCGCGCGCGTCCGCTGATCAGCAAG	28563	Db	29425	-CGCGCTCACCGATGCGCGCGCTCACCGCC-----TTGCTCCAGCCCC	29471
Qy	869	luAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluG	889	Qy	1190	nGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLy	1210
Db	28564	CTGCGGTGCGCTCCCGCACCGGTACCGCGTTCGCGCGCTCCGTCACCGCGCGGG	28623	Db	29472	ACCGGCCCC-GCCATCACCGGAGAGCTTCGCGACGCGTTCGCGCGCGCGCGCGCCAC	29530
Qy	889	lyAlaLeuLysAlaGluLysGluGlyGlySerGlyArgAlaThr-----T	905	Qy	1210	sGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleTh	1230
Db	28624	GCGCC-----GAAGCGGCTTCGCGCGCGCGCGCTGTGTGGAACCGAAA	28668	Db	29531	CGGCTCGCGGTCCCGAATAGCCCGGCTTCGCGCGTCCACCGCTGGCGCGCGTGC	29590
Qy	905	hrAlaLysSerSerGlyAlaProGlnAspSerSerSerAlaThrCysSerAlaAspG	925	Qy	1230	rHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleGlyGluAs	1250
Db	28669	CGGCTGTTCGCGGGCGGCGGCCCCCGCGTCCGCGCGGTCACTCCGTCGCGCCA	28728	Db	29591	CGCGCGCGCGCGCGC-----CTTGGCACCGTTCGCGCGCGCGCGCGCGCGCGCG	29621
Qy	925	luValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSer-ProArgProSerLeu	944	Qy	1250	pSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTy	1270
Db	28729	GGCTC---TCCGTCAAAATCCGGAGATAGTAAGCGTCACCGCCCAAGCCACCATTA	28785	Db	29622	-GCCAGCGCGCTCGCGCGCGGTGAGTGGCGCGGTGTTGCTGGCGTCCACCGCGCTG	29678
Qy	945	LeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLys	964	Qy	1270	rGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSe	1290
Db	28786	CCAGCGTCCCGCGCGCGCGCGTCCCGCGTCCCGCGCGCGCGCGCGCGCGCGCG	28836	Db	29679	----GAGCGGAGGTGCGGTGCCGTGTCCCGTGTCCACCCCCCAGACCGCGCGCGCTGC	29734
Qy	965	GlnLeuLysGlnArgAlaAlaIleProProIleGlnValThrLysValHisGluPro	984	Qy	1290	rLysGluAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgTh	1310
Db	28837	TTGCCGACCTCGCGGGTGGCGCGCG-----GCCCGC	28875	Db	29735	C-----GGGGTTCACCGCGGTTCACCGCGCGCTC	29761
Qy	985	ProArgGluAspAlaAlaProThrLysProAlaProProAlaProProProGlnAsn	1004	Qy	1310	rTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLe	1330
Db				Db	29762	CGCGCGCGCGCGCGCGCGGTGATACCAATGCGCGCGGT-----GCCGCGCGCGCGCGCGCGCGCG	29818


```

Db      31368  GGCACACCGTCCCCGCCA-----ACACCTGG----- 31395
QY      2035  uLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluPr 2055
Db      31396  -CTGCCCGCCGA-----CCGCGAGGTGGCTCAC 31424
QY      2055  oValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGl 2075
Db      31425  GCCAGCCCAACCGGCCCAACCGTCCAGCGCGGACCGCGGCTCCGCCATCA-----CC 31478
QY      2075  uGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProVally 2095
Db      31479  ACCGTTCGCCCTCAGATACGACGACGATTAACCGTGGAGCTCCGTACCA----- 31533
QY      2095  sLeuGlyGlyGluAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSe 2115
Db      31534  -----CCGGCCCGCGCGCCACCGCTTGGCGCGGACCGCCCGCGCGCCGCATC 31583
QY      2115  r-----SerSerProLeuLeuGlnThrAl 2123
Db      31584  GCCGGGTGGGCCCAACCGCGCGCGGCCCGCCACCGCGGCCCGCGTTCCTCCATCCTC 31643
QY      2123  aProGlyValLysGlyHisGlnArgValThrLeu-----AlaGlnHisIleSerGl 2141
Db      31644  ACCGGGGTACCGGATCAACCCAGGAGATCGCGTCATATCGCATAGCGCGCACCCGC 31703
QY      2141  uValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuPro-- 2160
Db      31704  GGCTCCGTACCGCCATAGCCGCGCGGCCCGCCCGTCAACGACCAACCGCGCTTGCAC 31763
QY      2161  ----AlaProLeuTyrSerPheProGlyAlaSer-----CysProValLe 2174
Db      31764  GGCACCGCGCTTACCGCCATGACCGCGCGGACCGGTGGCGGCGCATTCACCTTT 31823
QY      2174  uAspLeuArgArgProProSerAspLeuTyrLeuPro-ProProAspHisGlyAlaProA 2194
Db      31824  GCCCGCTTCCGCGCGTACCGCGCGCGTGTGGCGCGCGCGACCAATTG-GCCCGCG 31882
QY      2194  laArgLysProHisSerGluGlyGlyLysArgSerProGluProAsnLysThrServ 2214
Db      31883  CGCGAGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 31935
QY      2214  alLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluP 2234
Db      31936  -----CCGGGTGCGCACCGTGCACCGTGACCGCGCGCGCGCGCGCGCGCG 31966
QY      2234  roGlyHisSerArgSerAlaValTyrProLeuLeuTyrArGAspGlyGluGlnThrGluP 2254
Db      31967  CGTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 32011
QY      2254  roSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerL 2274
Db      32012  CGTCACCGCGCGCTCCACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 32071
QY      2274  ysLeuThrGluSerAsnSerAla-MetValLysSerLysLysGlnGluIleAsnLysLys 2293
Db      32072  CGACATACCCCAATAGTCCAGCCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 32131
QY      2294  LeuAsnThrHis-----AsnArgAsnGluProGluTyrAsn 2305
Db      32132  TTCGCGCGGACCGCCATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 32191
QY      2306  lIleSerGlnProGlyThrGluIlePhe-AsnMetProAlaIleThrGlyThrGlyLeuMe 2325
Db      32192  -----CCGGGCCATTTATTCGGTAGCCCGCGCGCGCGCGCGCGCGCGCGCG 32239
QY      2325  tThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaI 2345
Db      32240  CACCGGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 32294
QY      2345  eIleArgLysAlaLeuMetGlyLysTyrAspGlnThrGluLysThrProProLeuSerAl 2365

```

```

Db      32295  -----GAGAGCCACCCCTCTCCCGACCCAC 32317
QY      2365  aAsnAlaPheAsnProLeu---AsnAlaSerAlaSerLeuProAlaAlaMetProIleTh 2384
Db      32318  CGTTCCGCGCCACCCACCGGACCGCATCGCCCGCTCGACCTGCCAACCGTCCAC 32377
QY      2384  rAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLy 2404
Db      32378  CCCTGGC-----ACCGTCGTCGACCGCGGAACACCGACCGCGCCCAT 32419
QY      2404  sValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGl 2424
Db      32420  CCTGACCGCGGTACACCGTTCGGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 32479
QY      2424  y-----AspArgProProSerValSerSerValHisSerGluGlyAspCysAsnArgAr 2442
Db      32480  CCCCATGACCAACCAACCGCGGTACCGCC---CGCGCGCGCGCGCGCGCGCGCGCG 32536
QY      2442  gThrProLeuThrAsnArgVal-TripGluAspArgProSerSerAlaGlySerThrProp 2462
Db      32537  CACCGGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 32596
QY      2462  hePro-----TyrAsnProLeuI 2468
Db      32597  CGCCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 32656
QY      2468  leMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlys 2488
Db      32657  TCCGCGGTACTCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 32716
QY      2488  erGlyProLeuAlaGlyPro 2494
Db      32717  ATAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 32736

RESULT 91
AY211023/c 96742 bp DNA linear BCT 03-MAR-2003
LOCUS Streptomyces violaceoruber strain SANK95570 plasmid pSV2, complete
DEFINITION
ACCESSION AY211023 AH011227 AY055227 AY055228
VERSION AY211023.1 GI:28797237
KEYWORDS
SOURCE Streptomyces violaceoruber
ORGANISM Streptomyces violaceoruber
REFERENCE 1 (bases 1 to 96742)
AUTHORS Spatz,K., Kohn,H. and Redenbach,M.
TITLE Characterization of the Streptomyces violaceoruber SANK95570
JOURNAL FEMS Microbiol. Lett. 213 (1), 87-92 (2002)
MEDLINE 22123362
PUBMED 12127493
REFERENCE 2 (bases 1 to 96742)
AUTHORS Spatz,K., Scholz,C.J. and Redenbach,M.
TITLE Complete nucleotide sequence of the linear plasmid pSV2 from Streptomyces violaceoruber SANK95570
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 96742)
AUTHORS Spatz,K. and Redenbach,M.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2001) Department of Genetics, Genome Research Unit, Kaiserslautern University, Erwin-Schroedinger-Str., Kaiserslautern 67663, Germany
REFERENCE 4 (bases 1 to 96742)
AUTHORS Spatz,K., Scholz,C.J. and Redenbach,M.
TITLE Direct Submission
JOURNAL Submitted (18-DEC-2002) Genome Research Unit, University of Kaiserslautern, Erwin Schroedinger Str. 24, Kaiserslautern, Rheinland-Pfalz 67663, Germany
REMARK Nucleotide sequence updated by submitter
COMMENT On or before Mar 3, 2003 this sequence version replaced gi:22096410, gi:22096411, gi:22096409.

```

FEATURES
 source
 location/Qualifiers
 1..36742
 /organism="Streptomyces violaceoruber"
 /mol_type="genomic DNA"
 /strain="SANK95570"
 /db_xref="taxon:1935"
 /plasmid="pSV2"
 655..1137
 /gene="pSV2.01"
 655..1137
 /gene="pSV2.01"
 /note="similar to hypothetical protein from Streptomyces coelicolor A3(2) encoded by GenBank Accession Number AL939132"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAO50085.1"
 /db_xref="GI:28797238"
 /translation="MELDGMAPRGRGVFGDWMDEGVTTVSEALVGIADRWRRGGR SAGAELAVHTADVQRGAGLVLEFVRGHVLPTRPHTTARHVRLEALGTGLC TARTNADGVVCTLDAGHYDADDKPPFKDKEGGMKADASIWNLDGAACIPHAAL"
 1233..3836
 /gene="pSV2.02"
 1233..3836
 /gene="pSV2.02"
 /note="similar to Streptomyces lividans helicase-like protein and Streptomyces coelicolor A3(2) putative helicase encoded by GenBank Accession Numbers AF194023 and AL590463, respectively; DEAD-like helicase superfamily; helicase associated domain"
 /codon_start=1
 /transl_table=11
 /product="putative helicase-like protein"
 /protein_id="AAO50086.1"
 /db_xref="GI:28797239"
 /translation="MGPPARSSVPPQVGARS CORLARARRSPACALSFVDGRILV TVPTDLDAQTQAWRAVGHSSPMVAVCSLENDPMLNQLGVRTTNPICLAWAGHP VVVFATYASVDREVDPAEQCRKGRPLENALAGGERLIGQRMDFDLAIVDEAHGT AGDGRWAAIHDNRAIPADPLRYLTATPRILAAAPQKAGDOEAETIAMDPPNGT YGAWALEGTSEALEREITLAGEFIDVLEIRDPSPVGETEARGRERLALLOTALLIEF AAAYSLRTVMTFTHQKEAAAFADKLPTAAELVYNDASDADLAADAKLPASSIDAEF VELEGRRVPPDRVWSAWLGDGTVAAERREKIROFANGINAAGHRVRAFLASVRVLG EGVDITGRGVGEVASFATRCGSQVEIIVQNGIRALRLNRDGSCTKVARIIVPVEPNED PADWVASFPLVAVLQGLKSHDERLVEQLASALTSGRKKVHQWDEGRIVAGGG ESDGDEQOEQDGDTHDAAEASALLHFFSPSRDVATIAAFLRTRVREPSVWLEGYQ ALLRWBAEGITGVHAYTEVGVYTKDFPLGRVWQQKALRAGELEDREKRLTLLD APAGWMPGEAEWAENKLAALRRYRANGALPRQDAVWGEAEVVPQGMANLR KXGILCKDAERAARAQALAAVDEDNWCPMLDMORHYRVLADLVADGVLPAIEPGVL FEGDDIGRWLERQKPNQWALSITDQDERLTTLGTVTPRAPTKAGSKQGGKAQAFQ KGLAALAQYIAREGHRVPRAHAEIIVAEGERASVVVVLGVWVSNTKSRDKDKUTADQL AALAKLGVDA"
 3925..4764
 /gene="pSV2.03"
 3925..4764
 /gene="pSV2.03"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAO50087.1"
 /db_xref="GI:28797240"
 /translation="MAGTEPAPLWAGVDPDHLAVPVRQWLYGVLRDYSIAARVAVWL KLPISLTDQPSATLAAFEDETPMLRLIIDIATLGLHRVLEIETHAHEISGIAAESV MELEBILHESGATISRDSGLSWRINETLHYDKAVAGASGLMAQADHRLRAAFS RAYGFKPPDSAYSRAIKSVAEASPLFLNAPRPTTKGVRSGLDQGRKHVEMVIADK TGAPASIDAVVAMISLLWHGQDRDRHEGPTSPVQTQAEATAVHTAAILIHWISNGSI QKK"
 4815..5306
 /gene="pSV2.04"
 4815..5306
 /gene="pSV2.04"
 /note="similar to hypothetical protein from Streptomyces coelicolor A3(2) encoded by GenBank Accession Number AL939104"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAO50088.1"
 /db_xref="GI:28797241"
 /translation="MORGRGPOMGVPCCLATTSEGHFSQYGSASFGLSPGCGAGLT ALKQWIREGAGRPVRCGHTTEITVYSEAEFVVLKLGWISNFKPGAGLPRNSWTHC GARGHLRGIRGHYGASGAPRRPARRGPQTWHPARAASAGMIRAVRTVDRSYGALLCV CRR"
 5429..6127
 /gene="pSV2.05"
 5429..6127
 /gene="pSV2.05"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAO50089.1"
 /db_xref="GI:28797242"
 /translation="MRIRASRWQDAVRMLLLNAAAHFRPGVGDGEVPAECVGVVRA QVRQLKDFWRYPDYLAYELMTEYKAPDEPALLDAEILSSQDFLRFRPMLRHR FGAEELDNALAVLVERGLLKTQVLRKRVYVLLQAGRTLAVNIWVEAPALAA VVERTKRVVALVDGLGCTQLKKRQYLLRDYADTPIGSIYITPIDQARARLARALATAA PQAASAPADQGEAS"
 6124..8130
 /gene="pSV2.06"
 6124..8130
 /gene="pSV2.06"
 /note="similar to putative large ATP-binding protein from Streptomyces coelicolor A3(2) encoded by GenBank Accession Number AL939116"
 /codon_start=1
 /transl_table=11
 /product="putative large ATP-binding protein"
 /protein_id="AAO50090.1"
 /db_xref="GI:28797243"
 /translation="MTKPPDPVATPPVVPVPTADDILTQAAARLDVAKHQVERALAEANIA LTTTPAGDRRLRLRVDGIRKIDGEPFTDIRPFPGIWIIVHPENSAGKTSLLLEFLV WPLRAGDRLLPVGVRSMVRTLHLDLVVAGKPTIALEQIPGLLPVSVVDILTADTVQA LLAADNDPLRLVAKSGSDTDERVIGSPFLDALRMERTLSWSATCGADGEGAPQVHG WPAFGACIYNAGDELLLDGVSAPGLPAKLLERFVDIPISSAHQAQLSVAVKGRKTYT RQARCAEGLDHARSARAQWQALMQOEIOELSGAPETDIAPLQKADQAQELHL RRLNSLNGATCSLADAAARVAGBAQDLQAGETQWARRVGLRNLNPSCCPRCEPVGDE RRAARERATCAVCTRPLPEIAEDIAEAVLEQTTDIATALEAAHAAHRAVQASA DAATVAHQLEATVKSADYLFPTMDSEIVLSQFGVQHLNSVTLKRNQNVNVPVKEG VKHPFKLSRGDRLRMRVATVIALLRAGARHIDSHGILLIDSVAAEEMTADAAARL IDEIQTITRSLPELQVFTTQAQPELVEGLLPEDHIIITSSEHLF"
 8299..10257
 /gene="pSV2.07"
 8299..10257
 /gene="pSV2.07"
 /note="similar to hypothetical protein from Streptomyces coelicolor A3(2) encoded by GenBank Accession Number AL939116"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAO50091.1"
 /db_xref="GI:28797244"
 /translation="MAGPQLPYLLTALAAGHIDTIPAEQARALLAAATQGLRQPHSAV VLADALVTCN="MAGPQLPYLLTALAAGHIDTIPAEQARALLAAATQGLRQPHSAV GSPHRLALATEITGTEPADLDRLPLILIGIARDHFADTDLVLNALENQDLNPA TRADASELLADDERAALEADRRTTAGEQLQALMRFTELDRTHEARLDARAHAAIE AVLAFADLEQNPTGSARQLTEADAAHAAAMTGRWHQDLWISARSLTQSA EAPFVRNGLLHQALEAHADPQFTDHPDAOTLYDAVORROOLSSQAPQOQAMPQV

QY 1004 AsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSer-ProArgGlyLysse 1023
Db 51102 AAC-----CCCTCTTCGGCCCGCGTTCGAGTG 51073
QY 1023 rArgSerProAlaProProAlaAspLysGluAlaPheAlaGluAlaGln-LysLeuP 1043
Db 51072 CCCTTCGGCCCGCCCG-----CAGACGCTCC 51046
QY 1043 roGlyAspProPro-----CysTTPThrSerGlyL 1053
Db 51045 CCGCCACCCCGAGAGCTTTCACCCGCGAGCGACCGACCGACCCGATGACGTATCGC 50986
QY 1053 euProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspProS 1073
Db 50985 GGCCTTCGCCGAGCCG-----GTCTCGGCACCCCGTTCACCGCGCAATCCGA 50938
QY 1073 erAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaA 1093
Db 50937 GC-----CCGC 50932
QY 1093 rgProValLeuProArgProProThr----- 1101
Db 50931 GCGCGCGCCCGCGAGCCCGCGAGCGGGCGGTCTCGTGGCGGACCTTCGCTC 50872
QY 1102 -----IleSerAsnProPro-----ProLeuIleSerSerAlaL 1113
Db 50871 GGGGGGTGCGCAGCCCGCGCGCGCTGGTTCGCTCTCGTGGCGCGATCTCTCGC 50812
QY 1113 yHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnL 1133
Db 50811 GGTGTCTCTCGCAGCTA-----TGCGCGG 50788
QY 1133 euHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuP 1153
Db 50787 CATTCCTCCGCTCGCAGCGAGCC-----GGGCCAACCGGGAGCGCGGAGCGGTTCG 50731
QY 1153 roLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGlnLeuS 1173
Db 50730 GAAGACCGCGGCACCGCAAGAGCGGCCCGCGGGCGGTTCGAAGAAGCAGGTGCA 50671
QY 1173 erPro-----ArgGlyGlnAlaGlyProProGluSerLeuGlyV 1186
Db 50670 GCCCAGCCAGGTACGCGCGCAGAGCTT-CAGCGCGTCTCT-----GGCG 50624
QY 1186 alProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyV 1206
Db 50623 GCGCG-----GGAGCCCTCTCGCGCGCTCCGGTCCGCGCA 50588
QY 1206 lySerIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrA 1226
Db 50587 GGTACGGCGGGCGCGAGCGGTCTGTAC-----AGCGGTTCGGAGCGCGGGG 50537
QY 1226 rgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgI 1246
Db 50536 CGCGCGGTTCAGACCCGCTCGCCAGCGC-----TCCGCG----- 50499
QY 1246 leIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysG 1266
Db 50498 -----GGCGCGCGCGGTCTACGGGGCGCGCGCGGTTCGCGGCT---G 50447
QY 1266 lyHisValIleTyrGluGlyLysGlyHisValLeuSerTyrGluGlyGlyMetSerV 1286
Db 50446 GCGCTCCAACTCCGGGGGGGCAAG-----GGCGGTTCGGGG 50408
QY 1286 alThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGluThrAlaA 1306
Db 50407 TCACCCGCTCGGC-----GCACCTCGCAGCGCGCGAGCGTGAACCTCAAGAAGAAGCCGCG 50351
QY 1306 laProLysArgThrTyrAspMetMetGlu-GlyArgValGlyArgAlaIleSerSerAla 1325
Db 50350 CCGCGATCTCTCG-----GCGGCTCGCGGTTCGGAGGCTTCGCGCGCGGCA 50300
QY 1326 SerIleGluGlyLeuMetGlyArgAlaIleProProGlu----- 1338

Db 50299 GGGCCCGAGCAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 50240
QY 1339 -----ArgHisSerProHisHisLeuLysGluGlnHis 1349
Db 50239 CGGGTCCGGTCCCGCGGGGGAGACGACGACGACGACGACGACGACGACGACGACGACG 50180
QY 1350 HisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAsp 1369
Db 50179 CGGGCGCGGGG-----CGGCGG 50162
QY 1370 TyrLeuArgGluAlaLysLeuLeuLys-----ArgGluGly 1382
Db 50161 TTCTCTCGCGGTTCGGCGAGGAAACGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 50102
QY 1383 ThrProProProPro-----ProSerArgAspLeuThrGluAlaTyrLysThrGln 1400
Db 50101 GA-CCGGCCCCGCGACCAAGAGCCCCGCGCAAGGGCGACGGGGTTCGGAGGAAGCTTCG 50043
QY 1401 AlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGlu 1420
Db 50042 CCGGCACCCAGCTCGAAGGGCCCCGCGCACCCGGAAGGGCAGCAGCAGCGCGCGGAGG 49983
QY 1421 AlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuPro 1440
Db 49982 GCGGACGAGA-----AGTCGCGCGCGGTACGACC-GGCGCGTCCGGGATC 49936
QY 1441 LeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyr--- 1459
Db 49935 CTGCGCAAGAAGCCC-----GGCACCCACCTCTCCGGCACCGCGCGCGCGCGCGAG 49885
QY 1460 -----AspThrGlyAlaSerThr-----Gly 1467
Db 49884 CCGAAGGGGAGGCGACGAGCGCGCGCGCGCACCGCGCGCGCGCGCGCGCGCGCGCG 49825
QY 1468 SerLysLysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProVal 1487
Db 49824 GGAAGAAGGCGAAGACCGCTCCGCGAAGGGCGGACCGCTCGGTCGCGTCCGCGCAAG 49765
QY 1488 HisProLeuAspValMetAlaAspAlaArgAlaLeuGluAlaCysTyrGluGluSer 1507
Db 49764 GCGCGCGGGGGAAGTCCCGCAAGGCAAAAGATACGCTCAGCGCGGGGATCGCGCACG 49705
QY 1508 LeuLysSerArgProGlyThrAlaSer----- 1516
Db 49704 GCAGGTTCGGCGCGCGACGACAGTACGGTTCGCTGCCCAAGTCCGGCGCTTCCAGAA 49645
QY 1517 -----SerSerGlyGlySerIleAlaArg----- 1524
Db 49644 GGAGTACAGCCGCAAGTCTCGGGCGGATGAAAGTCCGACGCGCGCGCGCGCGCGCG 49585
QY 1525 -----GlyAlaProValIleValProGluLeu 1533
Db 49584 TCGCGAGCGCGGTTCACGAGACCGCGGAGAAGCGGTCAAGCGCGCGCGCGCGCGCG 49525
QY 1534 GlyLysPro-----ArgGlnSerProLeuThrTyr 1543
Db 49524 GCGCGCGCTCTCTCACCAGGTCAAGGTCAAGAGCGGAGAGAGAGAGAGAGAGAGAG 49465
QY 1544 -GluAspHis-----GlyAlaProPheAlaGlyHisLeuProArgGlySe 1558
Db 49464 CCAAGACCAAGCGCGCGGTTCGCGCGCGAAGCGCGGTTCGCGCGCGCGCGCGCGGG 49405
QY 1558 rProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerLy 1578
Db 49404 CACCGA-CCGCGCGCAGCGCTCTGTCCAGCGCAAGCGCGCGCGCGCGCGCGCGCG 49346
QY 1578 sAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSe 1598
Db 49345 ATCGGCACCAACCGTTCGCGCGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCG 49286
QY 1598 rThrValProGluHisHisProHisProIleSerProTyrGluHisLeu----- 1614

Db 49285 CCGGTTCCCG-----CCGCTCCGCCACCCCGGACCACTGCGCCCGCGCTCG 49241
Qy 1615 -----LeuArglyValSerGlyValAspLeuTyArgse 1626
Db 49240 GCCGCGCGGTTCACGGGAGTGCCTACTGCCCGCGCCCGCGG-----CGCGG 49193
Qy 1626 rHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAsp----- 1644
Db 49192 GCACGTCCG-----GGCACCGGCACCCCGCGCTGTCTCGCGGTTCGAGCGGAT 49145
Qy 1645 -----AlaAlaAlaAlaTy-TyrLeuProArgHis-----LeuAlaProAsp 1659
Db 49144 GGGCATGCGAGCGCGCGCTCCGCTCCGCGCGCCCGCATCAGGTGGAGCGCGCGACTC 49085
Qy 1659 roThrTyProHisLeuTyProProTyTyLeuIleArgGlyTyPro--AspThrAlaAl 1678
Db 49084 CCGGCATCCCGCGCGCGAGTCCG-----GCACGGGGTTCCTCCGCGCGCACTCGAG 49031
Qy 1678 aLeuGluAsnArgGlnThrIleIleAsnAsp----- 1688
Db 49030 TAGAGCACGAAGGAGAACAGCGTGTAGTGCCTGCTGACGAAGTGGGCGAGCGGAG 48971
Qy 1688 ----- 1688
Db 48970 GTGCGCGGACATGACGCCCGAGACGAGACGAGTCACTCGACGACACGCTCGAGCT 48911
Qy 1689 -----TyIleThrSerGlnGlnMetHis---HisAsnThrAlaThrAlaMetAlaG 1705
Db 48910 CTTGGAGGACTCACCGCGAGTCTTCAAGGCACACGACCTGCAACAGTGGCGCG 48851
Qy 1705 nArgAla---AspMet--LeuArgGlyLeuSerProArgGlu----- 1717
Db 48850 GAAGGCGCGGACCTCGCTACGAGTGCAGCAGCTCGCGCAGACCTGCGCTCCACCCA 48791
Qy 1718 -----SerSerLeuAlaLeuAsnTyAla 1725
Db 48790 CAACGTATCGCGCGCTGACCGCGCGCGATGGCCAGCTGGCGAATCCATGGAGCT 48731
Qy 1726 AlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValPro 1745
Db 48730 GTTGGCCCGAAGCCACAGATGGCCACCAAGTCTCTACCG-CCGACAGCTCTCCG 48672
Qy 1746 ProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyLeuProThrAlaPro 1765
Db 48671 AGAGCGCGGAACCGCATGTTCAGC-----CCTACCGGCCCC 48633
Qy 1766 GlnProPheSerSerArgHisSerSerProLeuSerProGlyGlyProThrHisLeu 1785
Db 48632 -----TCCAGCAAGCCACCGAGGACCGCGCTGGCGCTGCCCGCG----- 48591
Qy 1786 ThrLysProThrThrThrSerSerSerGluArgGluArgAspArgAspArgGluArgAsp 1805
Db 48590 -----CCCGCTCCACACGAGNACTGAGAGGAGACACGACGTGAGCAGCAACCTCCC 48537
Qy 1806 ArgAspArgGluArgGluLysSerIleLeuThrSerThrThrValGluHisAlaPro 1825
Db 48536 CGC----- 48534
Qy 1826 IleTrpArgProGlyThrGluGlnSerSerGly-Ser-----SerGlySerSerG 1842
Db 48533 -----CATCCCGGGCATGAACAACCGCGTGGCGAGCAGCAGCCGAGCGCCAGCGGG 48480
Qy 1842 yGlyGly-----GlyGlySerSerSerArgProAlaSer----- 1853
Db 48479 CAACGGCTTCTCGCTCGCGCGCGGACCGGACGCTCGCGCTCCCGCTCTTCCT 48420
Qy 1854 -----HisSerHisAlaHisGlnHisSerProIleSerProAr 1866
Db 48419 CAAGGAGGCGATGCACCTGCTCCGATCCACATCGAGCGCCACACGCCAGCGCGCCATCG 48360
Qy 1866 gThrGlnAspAlaLeuGlnArgProSerValLeuHisAsnThrGlyMetLysGlyII 1886
Db 48359 CTTGTCGGAGATGTGGCGGCAG----- 48338

Qy 1886 eIleThrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSer-Serp 1906
Db 48337 -----GCCGAAGTCGAGCCCGCTTCACTCGATGATCGAGGACACCTCCACCGCCCTCAC 48282
Qy 1906 roVal-----ArgProAla-----AlaThrPhePro---ProAlaThrHisCysp 1920
Db 48281 CGGCGTGGCGAGCGGTCCGCGGAGATGTGTCAGCGCGCGCGCGCTCGACACCGGCGC 48222
Qy 1920 roLeuGlyGlyThrLeuAspGlyValTyProThrLeuMetGluProValLeuProL 1940
Db 48221 CCGCGGCTTCAAGGTGCGCACACCGGGAGTACC-----GCCCGGCTACGAGCGCTC 48168
Qy 1940 ysGluAlaProArgValAlaAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheL 1960
Db 48167 CAACGGCTCCGGAGTCCAGCAGCGCAAGC-----CCGGCTTCTTACCGCG----- 48124
Qy 1960 euAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerG 1980
Db 48123 -----GCACCGCGCGCGAGCTGAACAACCCCTCACCCACACGAGCAGCGGCGCG 48072
Qy 1980 luProArgProLeuValProValSerGlyHisAlaThrIleAlaArgThrProAlaL 2000
Db 48071 CACCGGAGCG-----CCGGGTGCGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 48027
Qy 2000 ysAsnLeu-----AlaProHisHisAlaSerProAspProProAlaProAlaSer- 2017
Db 48026 GGAGGAGTGTTCGAGCACCATCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 47967
Qy 2018 -----AlaSerAspProHisArgGluLysThrGlnSerLysp 2030
Db 47966 CCGCGGCGACGTCACTCACCGCGCGCTCTGCGCGAAGCACCTCTACGCGCTCG----- 47911
Qy 2030 roPheSerIleGlnGluLeuLeuArgSerLeuGlyTyHisGlySerSerTySerP 2050
Db 47910 -----CGCGCGCGGAGTGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47877
Qy 2050 roGluGlyValGluProValSerPro-----ValSerSerProSerL 2064
Db 47876 CGACAGCGCTGTCTCTACTCTCGCTCCGCGGAGTGGACTGGGACACCTCGCTGCTGCT 47817
Qy 2064 euThr-HisAspLysGlyLeuProLysHisLeuGluLeuLeuAsp----- 2078
Db 47816 GTACGCGAAGACCAAGACCCCGACGAGGACTGAGCGTCTGACCCCGCGCGCGCGCGCG 47757
Qy 2079 -----LysSerHis 2081
Db 47756 GATCCCGCTTCGACGCGCGCGCACCTTACGTGGAACTGATCACTCCCGCGCTTCAA 47697
Qy 2082 LeuGluGlyGluLeuArgProLysGlnProGly---ProValLysLeuGlyGlyGluAla 2100
Db 47696 CTGGTGGAGTCTGCGCCCGCGCGCGGTGGGTCTCTGATGGGATGGATGGC---GCC 47640
Qy 2101 AlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerProLeuLeu 2120
Db 47639 GCTCAC---CCGACGCGCGCGCT-----GGTCTCTGGAGCT 47607
Qy 2121 GlnThrAlaProGlyValLysGlyHisGlnArg-----ValValThrLeuAlaGlnHis 2138
Db 47606 CACCGAGCACCGAGCGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 47547
Qy 2139 IleSerGluValIleThrGlnAspTyThrArgHisProGlnGlnLeuSerAlaPro 2158
Db 47546 CTTGCGCGGAGTGGCAGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47487
Qy 2159 LeuProAlaProLeu-----TyrSerPheProGlyAlaSer 2170
Db 47486 GTTCCGCG-CCGACCGCGCGACTTCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 47428
Qy 2171 CysProValLeuAspLeuArgProProSerAspLeuTyLeuPro----- 2186
Db 47427 TGCCAC--CTTACCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47370

2187 QY -----ProProAspHisGlyAlaProAlaAtgGlySer-----ProHis 2199
|||||
47369 Db CAAGTCGCGCCGATGAAGGGCTCCGGGCGCGCGGATCGCGTGGTCCGCGCCAC 47310
|||||
2200 QY -----GlyLysArgSerProGluProAsnLysThrSerValLeu 2215
|||
47309 Db CTCGAGGACGTCGACCGCGCGGAGGATCCGCGGCT----- 47268
|||
2216 QY GlyGlyGluAspGlyLeuGluProValSerProProGluGlyMetThrGluProGly 2235
|||
47267 Db ---GTGGGAGACTCGGCTCTGCGC-----CCGGGCAAGCGCGCACCC 47229
|||
2236 QY HisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSer 2255
|||
47228 Db CGGTGTGAGCTGCTGACTACCG-----GGCGAACCAGCCGCGCAT 47187
|||
2256 QY ArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPheSerLysLeu 2275
|||
47186 Db CGTGCTGCTGCTCGCTCCCTCCCGCAAGAGATCCACTCCGCGCACCGCGCATCTG 47127
|||
2276 QY ThrGluSerAsn-----SerAlaMetValLysSerLysLysGlnGluLeu 2290
|||
47126 Db CTCGCGCTTCGGCATCGACGCCCTCCGGCTCGTATCGAGACCGCGCATCCGCCA 47067
|||
2291 QY AsnLysLysLeuAsnThrHisAsnAtgAsnGluProGlu-TyrAsnLysSerGlnProGlu 2310
|||
47066 Db GGGCGTCTCTC-----CATCTACAAGCAACCCGCTGATGAGGTCCGCAAGCCAC 47013
|||
2310 QY yThrGluLeuPheAsnMetProAlaLeuThrGlyThrGlyLeuMetThrTyr-ArgSerG 2330
|||
47012 Db CGTCGAGACCTGACGATGACGCC-----AAGGGCGCGCTCACCATCGCGCTCTG 46962
|||
2330 QY lnaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaLeuArgLysAlaL 2350
|||
46961 Db CCACGACGCAAGCCCGCGGATGCGGCTGTGGACCCGAGCGCGCGCCCTGCGCG- 46903
|||
2350 QY euMetGlyLysTyrAspGlnTyrGluGluSerProProLeuSerAlaAsnAlaPheAsnP 2370
|||
46902 Db -----GCATCACCGCGCGGCTCATCCGCGCGCGGCAAGTC 46869
|||
2370 QY ro-----LeuAsnAlaSerAlaSerLeuProAlaAlaMetProLeuThrA 2385
|||
46868 Db CGTCTCGAGAACCTGATCTCTCGCGCGCGGAGAGCGCTCGCGGCTCTCTCTGGTCTCG 46809
|||
2385 QY laAlaAsp-GlyArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLys 2404
|||
46808 Db CGAGCTCAGGCGCGCATGAG-----CCTCCGAGGCGGAGG-GCCGGG 46765
|||
2405 QY ValSerGlyArgProSerSerArg-----LysAlaLys 2415
|||
46764 Db TCGACTGTTCCCAAGGCGCGGTGGAGAGCGCTCTGATGCTCACGCGCGCGCACCGG 46705
|||
2416 QY SerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSer----- 2432
|||
46704 Db TCATGAAGTACCGCGAGCGCATCAACGACGAGATGGCGCGCGGACTTCGCCCTGAACA 46645
|||
2433 QY -----ValHisSerGluGlyAspCysAsnArgArg 2442
|||
46644 Db AGCGGTGCGCGCTCATCAACATCACTCGGTGAGAGATCAACCGGCTGTCTCCACCGG 46585
|||
2443 QY Thr-----ProLeuThrAsnArgValTyrGluAspArgProSer--- 2455
|||
46584 Db ACGAAGCGATCGCAAGGCCACCGCTACCTCATCGCGGACATCCAGAACCGCGCGCA 46525
|||
2456 QY ---SerAlaGlySerThrProPheProTyrAsnProLeuLeuMetArg----- 2470
|||
46524 Db AGGTGCGCATCGCATCCGCTCGCGCTGCGAGTCCCTCACCTCAAGGACCTCGGTGAGC 46465
|||
2471 QY -----LeuGlnAlaGlyValMetAlaSerProPro 2480
|||
46464 Db ACGACCGCATCGCGCAGGCGCAAGGTCCGGCGGCTGTTCTCATCGCGCACCGCTCCT 46405
|||
2481 QY ProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHis---AlaTyrAsp 2499
|||

46404 Db CCTCCA-----CCAAAGGAGATGGCGCTGGAGC 46378
|||||
2500 QY GluGluProLysProLeuLeuCysSerGlnTyrGluThr 2512
|||||
46377 Db GCATCGCCCGCGCGGTT-----TCCAGATGGAGAACA 46345
|||||
RESULT 92
AY310323/c 138203 bp DNA linear BCT 06-DEC-2003
LOCUS Streptomyces sp. FR-008 heptaene macrolide complex synthesis gene
DEFINITION cluster, complete sequence.
ACCESSION AY310323
VERSION AY310323.1 GI:34766435
KEYWORDS Streptomyces sp. FR-008
SOURCE Streptomyces sp. FR-008
ORGANISM Streptomyces sp. FR-008
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 138203)
AUTHORS Chen, S., Huang, X., Zhou, X., Bai, L., He, J., Jeong, K.J., Lee, S.Y. and
Deng, Z.
TITLE Organizational and Mutational Analysis of a Complete
FR-008/Candidicin Gene Cluster Encoding a Structurally Related
Polyene Complex
JOURNAL Chem. Biol. 10 (11), 1065-1076 (2003)
PUBMED 14652074
REFERENCE 2 (bases 1 to 138203)
AUTHORS Chen, S., Huang, X., Zhou, X., He, J., Bai, L., Jeong, K.J., Lee, S.Y. and
Deng, Z.
TITLE Direct Submission
SUBMITTED (30-MAY-2003) Bio-X Life Science Research Center,
Shanghai Jiaotong University, 1954 Huashan Road, Shanghai 200030,
China
FEATURES
source
1. .138203
/organism="Streptomyces sp. FR-008"
/mol_type="genomic DNA"
/db_xref="taxon:206662"
complement(574. .1950)
gene
/gene="fscO"
complement(574. .1950)
CDS
/genes="fscO"
/notes="putative FAD-dependent monooxygenase"
/translation="MFSTCPAPTSIGFGCPHRRKGRTSRRIAVTPRCQKVRKGH
QAKWVRVMRGSSVAVVGGSIAGCAALAAASRGGAERTVFERADARLRDRGVIGV
HDARYAELRDAGYADMEPWAPLNRVWSVRDGAEGRAIGTQPPFPFRAYNWSLWS
ELRRVPETATYRAGAKBAVEQADGVTVRADGEPFDLVIGADGYRSVVREMY
PTGAAAYAGYIGWGTSPDVSGLPSDGLDAHNITPPGHCMAVRI PDGSGVSRHLNWL
YAPPRIDGHLDPRLPTSLPGRNLNAELTEHLRALVAEHPFPFWAKILSTPAETFL
IQYDLDLPHYATDRNLVIGDAASVARHPHGSLGSKALQATALAAEAAVAGESWEI
LEGYHAARGPVGVTAMVGLARRMGSAQVEETPDWSAMQAEFDAAWQEQNNQSDRRSGF
GHSLSKSR"
complement(2264. .3037)
gene
/gene="pabC"
complement(2264. .3037)
CDS
/genes="pabC"
/notes="putative 4-amino-4-deoxychorismate lyase"
/translation="MIELDGEAPAGPALASLALTNGHTTLLVNGRVRGLDIHLER
LIRDCTLFDALDPDRVKLARRAAPTGRATRVTVFDPALNLGNIAADARPGLV
TSRPAKPKPGLRVSVVHRRDLPEVKSGLCTLLRRLRQAQAGYDVLFTGPDGD
ILGGTWNVGLVRDGEVWPGGLVAGTTRQLLRGHRHSRTGRPRGRLSRGLRD
QRRGVGPVTGIDIDREFPAAHHSVTRLAETIYQALPGSPL"


```
gene
complement(3150. .3818)
/gene="fscRI"
CDS
complement(3150. .3818)
/gene="fscRI"
/notes="putative transcriptional activator"
/codon_start=1
/transl_table=11
/product="fscRI"
/protein_id="AAQ82551.1"
/db_xref="GI:34766438"
/translation="MDPAPADAGSPERAAVLADGPDRAAGAVACLDPSLTIQOVNQ
EFPRFGSSSELCGRSDLVHPSVQOPLMHQFARMKDGRRHFATEVIAVQERTFA
STPLNAGAVRGRTPDVAAILVMNNAEEEGADADVAPRKLLSEIDARILEGTA
GVSTIPLASRLYSROQVEYHVTGLLRKVPNRAALVSRAYSMGVLKVGTPPKVVE
DFIK"
complement(4377. .7205)
/gene="fscRII"
CDS
complement(4377. .7205)
/gene="fscRII"
/notes="putative transcriptional regulator"
/codon_start=1
/transl_table=11
/product="fscRII"
/protein_id="AAQ82552.1"
/db_xref="GI:34766439"
/translation="MLLERTTEIALVDAAALGAATEGRSSVLTLTGPIGIRGSAALLQRL
SCLADRDVRLRAHAAPMEQPFAGVVRQLPETLIGDSPEDARDRMDAHAFARH
VLADDAAPGACDAIAATEAVLHGLLSLLANVSADSRLLTLVDLWSDVPSRLWLT
LAKRLHGLAVVNCALDGDPRSHHTLAREIREAGTQLRPASLSLTATHELREHG
EAGDGFVQACHASVGNPLJLSILVGTGLGRRRPLAETHAETARRRPSQLRERAS
ILRTQPAVDLAAATAIAGESDAPTALRGLDSIGYAGALRALGALGALAAPDEP
RTHRSVRDAEESTLWQERWHEDEAALLYAAGSPAEQVAQLMAVTVTRQPAVD
VLAADATLRRGADPTAAGYLCRALDSPAAGVGRGLVELGTBERGDPDCLACERH
IAOAMTLPEPRRAVAALRISPTALGPA PLTAVDLLRQAAEDLGPADGILGTDRDLA
LRLEARLHCHGEDPHELASARHLKELPEPPVDGTGGERELIALLGAATLGCPLPA
SEVVRGLAVLDREPATSARHSTLPLAITMTIAADSVEAASVNLVAHEHANEGRITP
ANALVIEQALVHLGRQAQRESEALGTMTPEAHQGLDTAATMTHTVALESRDP
SLAEKVLRASTRIRPQGLTIGLRLLEAREIAAGDQGTALDLSHSCGRQLBAAGWH
NVLFPWRWAGVGRIGDIRTARSLAEBEYAEQWAGPVGVGRALRLGLLEDGG
RGIGLLREAVTVLGRSANOELARALSTLAEIRGGGAETELAREATLATSAGPL
AERAGAGHRSGPPATPEVLPFTERRVVVGLACRGLTNGEIAAGLVSSRAVEKHLT
HAYRAGISGRREILVLFSAEYVAGTD"
complement(7210. .10254)
/gene="fscRIII"
complement(7210. .10254)
/gene="fscRIII"
/notes="putative transcriptional regulator"
/codon_start=1
/transl_table=11
/product="fscRIII"
/protein_id="AAQ82553.1"
/db_xref="GI:34766440"
/translation="WYDLNMPPIELPILDRDLAALSAVIGELSGRPVAVTVTCEP
GLGNDLLRWAAATADAGRVLSAHATPAEHEVRYGVVAQLLAENRALAPRLFTD
EQCGGLPLGNLAVDRPTLLVVRDQWDLPSLRWLEALARRLPAPFALLTST
GTALTPEWVGWGLTGLATTIELALPPLTSGGTATVLRFAFGAGDPDPAFTALAEAT
RGIPAVVDLVRFAAGYSPRADRIDPLRATAEVVGDHATRALSDPDPAAVDLAR
ALGVDDLDPLVCTLAGPHSVSESRALAAASGITTLDGHPRVQDVAVRVARE
EMPAADBAELIARAAGLAQRAVADDDQIADLILLRVPVGPVAVDILRGFTSALRG
RDLAVAYLARLDEPAEDRAKIEFQLASVEMVTPAAERRLGSLIRATRPPTA
GLRARATDCLCGDTRAAHALAGIDSAPAEPRRGTGAPVPEEPAPSTPT
HLVHPRYATGPGATGDRGDAAPDLPGTPRAHPDAPRSPGPEAAGRPTAGDEDP
AVARHEGLTVLCVASFLGRDDADLVLPVPAHPDAVNPVAGVMAWHEGVAGVRE
EVRLEARSAPVATGRLVAVPRLIGCRALLTTDGDDEAENHGLMLDDESHERRAI
SVHILTVRAELHHRGPDAARDLAAQAEPLDLRLHPLFLPLMLALSMTDLQNG
HTHARERTARPLPLAHESATTAQLLFARGVLARTDDEPHQREHFRACGRWLLRHG
CANPAQVPMRSLAAEAHTLGTDEEARLVHEEVRLARWGAAPLGRAQLSLAVPE
ENVENLRAATVTLISAASATYTRAVIELAAAREENERRTAVLQAGPDTLPPVP
PGPHVPSAVARPGPPVATLGTGRRTAKAPCHDAAVMHGLTAVERTAAALAAQCLGNR
EIAATELVATTVRAVELRSLGVRKLRIRGRELRAALVQAEBS"
complement(10298. .13315)
/gene="fscRIV"
complement(10298. .13315)
/gene="fscRIV"
/notes="putative transcriptional regulator"
/codon_start=1
/transl_table=11
/product="fscRIV"
/protein_id="AAQ82554.1"
/db_xref="GI:34766441"
/translation="MMLWAGDRSGAAGPQRSGSAGPSLAPSDARESPVPHRSRVP
TLVSGSRVADVATGYSVRAALFGSLGTGRKGRTSPLLVGGARRALPALAADPCELOA
PAPPLVGRAHQOLDVLRHADARAARAGPRVLVLDGPAGIKGTALLRAALADGPELAGMT
DAGTSFVGLGLYLAANLMAADGPIVLVLDVHMCDESLRMLDFLLRADGELLVJV
AARTGTLPAPDALADLVNHLUPASLGLGLPLDTAEVAELTASFSFPQDPLPSFVRQ
LSVTGSGPLIEIVRLRLDRSAGLGGDDTGTARLAQAGKVVAAHVSFQBHPQVWVRQ
VARAVAVLDEDETRYLAAALGRVSVHVEAEIYLRDAGTLHPGHELSHDLVRAAVID
ACASGVAVLRRRAALLSDVGSPEEIAQTLVLPCTDDMAVSLVLRDAAAQAEQIG
ACEAAAYLERVREARPKDPLVSLRGLKLAETDPARSVTLTLEHSHSLTDDVRAAAT
AVOLGTCLAVQOQDGRALTEALDALDAELGPEPEPADRELTLTESALLIVGSD
KATLPIILRTEGLTPQDGTPAQROQLAMLVLSAEGGAEETVGRARRALRAGV
PLVGSLLPTSLALSADENAEAEVLETVLRSGSDTAAVNTVYVLAUSTRSLFRLENG
AVPDMAQAQTALEIILQOERWGTATMPHTYASVLTREGDPGRLEALDAIKRPHLD
RFVMEYHVMARGRALAADCDLDGALQVFGSCSMAEAGLTNPVLAPWMLTACILL
GEAGREEAGRAAAGHTRLAERWGTTRRQALGYAALARGAAAGTARTCALREAVALLAD
SPARSARGSLLLGRALVEDGAVRSRSHLREAVGLARCGCVALARARDELIAAG
GNRVTAPLMDLTGTERTVAGLVASGAGNREVAESLFTVTVTVELHLSVTRKLV
ARRGDTLEALREAGATAPAAARRPPGHAKRNP"
13522. .14898
/gene="fscMI"
13522. .14898
/gene="fscMI"
/notes="glycosyltransferase responsible for mycosamine
attachment"
/codon_start=1
/transl_table=11
/product="fscMI"
/protein_id="AAQ82555.1"
/db_xref="GI:34766442"
/translation="MDSAARPIFLVSLPESGLNPLLVLAGELARRGVPLDWFATDEH
RREERALSVDKSVSFASGEVYVPELSAVTWDDVREVTQSSRFKAHRAVVRQSYRP
ALQARKYRELEQVDEVRPALMVVDVAGVGVDLALARGIPYLVNVPFVASVNTLSH
PFGASVTPKSPVPNSGLPARMSVROKLANLTFKWTGLMFLHPDMAALREDAAIRK
ELGIAIPNATWRVDEAAVVCSSVAELDPFDPIDRVSLVGLVLPPLPEAPDDDEVR
WLDQASVVTWVFGTITRLTREVAALVEVARRMSGTHQFMKLPKQQLPBEAGSL
PDNLRVESVWPSQLDVLAHPNVFFSHGGNAYHEGVYFGKPVQVVRPLWVDCPDQAV
RGDRFGISLTLDKPTVDPDDVVDVVDKLTTRVTSDFRTEARLQALQRAAGRAAALD
VTGLLPAA"
14953. .16011
/gene="fscMII"
14953. .16011
/gene="fscMII"
/notes="putative aminotransferase responsible for
mycosamine biosynthesis"
/codon_start=1
/transl_table=11
/product="fscMII"
/db_xref="GI:34766443"
/translation="MAFTHPVSRPALDGRLELYVSDAVSGWISSQGPYVRRFEAFA
EWNGVAGVACSSSTAAALTLARALNIGPGEVIVPEFTWVASAWVYTGATPVFVD
CGDDLNDIVTRIEKITPRTRAMPVHVYGRCCMDVMDLALQYNLVVVDSAEAHG
VRVQDQAFSLFANKIITAGEGVCLTDDPRLAEQALHARAWAFTKRDHSLFKKLAY
NYRMTAQGVALAQTREDEILATREIAREYDAGLKDLPGITLMPARDVLMYDLR
AERRELRHLARGIETRLFFKPMRQPGYLDVPMVPTLNHARFSEDLGLYLPHTGLT
AAQDEYITGAVDRDFRAG"
Alignment Scores:
Pred. No.: 0.000152 Length: 138203
Score: 662.50 Matches: 651
Percent Similarity: 30.22% Conservative: 247
Best Local Similarity: 21.90% Mismatches: 1082
Query Match: 5.01% Indels: 1004
Gaps: 1
US-09-522-753-5 (1-2517) x AY310323 (1-138203)
```



```
QY 1227 ySerIleThrHisGlyThrProAlaaspValLeuTyrLysGlyThrIleThrArgIleIl 1247
Db 62954 ----- 62954
QY 1247 eGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHi 1267
Db 62953 -GGCCCGACGCCCGCACCGG----- 62933
QY 1267 sValIleTyrGluGlyLysGlyHisValLeuSerTyrGluGlyMetSerValTh 1287
Db 62932 -----GTCGC 62928
QY 1287 rGlnCysSerLysGluAspGlyArgSerSerGlyPro---ProHisGluThrAlaAl 1306
Db 62927 CTGGCTCCCGTGGACCGACACACCTCTCCCGCACCCCGACACCGTCCGCGGGC 62868
QY 1306 aProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlase 1326
Db 62867 C-----GTCATCGGTGCCCCGGACGGCTCACCGCCCTCGA 62829
QY 1326 rIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHisHisLeuLy 1346
Db 62828 ACGCAGCGGC---GCCAC-CGTCCAGCTCCACCCCGACCTGGCGGCCCTCACCCACGCCCC 62773
QY 1346 sGluGlnHisIleArgGlySerIleThr---GlnGlyIleProArgSerTyrValGl 1365
Db 62772 TCGACACGACGAGCGCGCACCTCCCGCCCTGTCGTCGCTTCCACGGCCA----- 62721
QY 1365 uAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLysArgGluGlyThrProPr 1385
Db 62720 -----CGACAGCG-GCAGAGACCCGACGAGGAGCGACACTCT 62681
QY 1385 oProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLe 1405
Db 62680 CGGCGCGGTGCTTAACGCTG-TCCACCGCTGGCTCCGGAGCGCGCCCTGGCCGACGC 62622
QY 1405 uLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaArgSerIl 1425
Db 62621 CCGGCTCTCTGTGTACCGAGGCGCGCTGCGCACCC-----GGCCCGGACGA 62574
QY 1425 eHisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgPr 1445
Db 62573 -CACCGT---CCCCGGCTCGCCGACGCCACCGC-----CTGGGGCTGTCTCC 62530
QY 1445 oLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerTh 1465
Db 62529 GCTCCGCCACACCGA----- 62514
QY 1465 rThrGlySerLysLysHisaspValArgSerLeuIleGlySerProGlyArgThrPhePr 1485
Db 62513 -----GCACCC---CGACCGCTTCC 62497
QY 1485 oProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCys----- 1503
Db 62496 TCCTGGCGCACT-----CGACCTGCCCGC 62470
QY 1504 ---TyrGluGluSerLeuLysSerArgProGlyThrAlaSerSerGlyGlySerIl 1522
Db 62469 CGCACACCGACGAGACAGCCCGCCCGCCCTGTATCGCGCGCGCGCGCGCGCGC 62410
QY 1522 eAlaArgGluAlaProValIleValProGluLeuGlyLys----- 1535
Db 62409 GCACCGCGGACACA-----GTTCCCGTACGGCGCGGACCGTCACCGTCCGCGCGCC 62356
QY 1536 ---ProArgGlnSerPro----- 1540
Db 62355 TGGTCCCGCGCGTACCCCGGAGGCGCGCCCGCGCGCGCGCACCGCTCCGTTGGGCGC 62296
QY 1541 -----LeuThrTyrGluAspHisGly----- 1547
Db 62295 AGGGAACCGCGCGCGCACCGCTGTGTATCAGGGCGGGAACCGGCACCTCGCGGCCCTCG 62236
```

```
1548 -AlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrPr 1567
62235 TCGCCCGTCACTCTGTACCCCGCAGCGGTACGCACCTGTGTCTCACCACCG---CC 62179
1567 oArgLeuGln-----GluGlySerLeuSerSerLysAl 1579
62178 GCGGCGCGCGCCCGCGAGCTGCGGCCTCGCGGAGAACTCGCGCGCTCGCGCGCG 62119
1579 aserGlnAspArgLysLeu---ThrSerThrProArgGluIleAlaLysSerProHis-- 1597
62118 AGGCAGAGATCTCGCTCGCAGCTGCGCAGCGCGAGCAACTGGCCGCGCTCCTCGACG 62059
1598 -----SerThrValProGlu---HisHisProHisProIleSerProTyrGluHi 1613
62058 CCGTACCGCGCCGAACACCCCTGACCGCGCTGTCACACGCGCGGTGTCT-- 62007
1613 sLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAs 1633
62006 -----GGAAGACGGCAC----- 61995
1633 pProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProAr 1653
61994 ----ACTGACGGCTCTCACCGCGACCGCTCGCCACGCTCTCGCGCCCAAGCGCAGC 61939
1653 gHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTy 1673
61938 CGGCTTGGCACCT-----CCACGAGCTCACCGCGCACCAACAACTCACCGCTT 61891
1673 rProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGl 1693
61890 TCGTGATGTCTCTCTACTACGGGGCGTCCG----- 61860
1693 nGlnMetHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLe 1713
61859 -----CGTGGCGCCCGCGCAGCGCACTACCGCGC--- 61830
1713 uSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAs 1733
61830 ----- 61830
1733 pLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlaTh 1753
61829 -CGCCAAACGCTTCTCTCGACGCGCTCGCGCAGCAGCGCGCGCGAG---CGGCTCTCCCG 61774
1753 xAlaMetAspArgLeuAlaTyrLeu-----ProThrAlaProGlnProPheSerSerAr 1771
61773 CGCACTCTCTCGCTGGGCGCTCTGGAGAGACCGCAGCGACCTGACCGCGCGCTCGACA 61714
1771 gHisSerSerSerProLeuSerProGly---GlyProThrHisLeuThrLysProThrThr 1790
61713 CCACCGGTCTGGCGCG-----CCTGGAACGCTCCGGCATCCGCCCTCACCGCGCGAC 61660
1791 ThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluArg 1810
61659 AGGGGTCTCGCTCTCTCGACGCGA----- 61636
1811 GluLysSerIleLeuThrSerThrThrValGluHisAlaProIleThrArgProGly 1830
61635 -----GCACCGCGCTGGACACCGCC 61615
1831 ThrGluGlnSerSerGlySerSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1850
61614 ACCTGTCCCGTACGCTCGACACCGCA----- 61585
1851 ProAlaSerHisSer-----HisAlaHisGlnHis-- 1860
61584 CCTGCGCGCGCAGAGGTGCCCTGCTGCGGCGCTGGCAGCGCGCGCTAGCAG 61525
1861 SerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsn 1880
61524 GCGCGCGGACTCGGGC-----CCGCGCGCGCACCGCGGCTCCGCGCG 61480
1881 ThrGlyMetLysGlyIleIleThrAlaValGluProSer-----LysProThrValLeu 1898
```


QY	2470	gLeuGlnAlaGlyValMetAlaSerProProGlyLeu-----	2484
Db	59446	-----AGGAGCGCACCCCGCGGACCGGTGACGCCACCGACCGG	59406
QY	2485	-----ProAlaGlySerGlyProLeuAlaGlyPro-----	2495
Db	59405	CGCGTACCCCTCGGTCTCTCCGCGGTTCGACAGCCGCTAGCGCCCGACGGCGGCA	59346
QY	2495	sHisAlaTrpAspGluGluPro	2502
Db	59345	CGACTCTGCGCCACCTCACC	59324
RESULT	93		
LOCUS	BHV1CGEN/c		
DEFINITION	Bovine herpesvirus type 1.1 complete genome.	linear	VRL 07-APR-2000
ACCESSION	AJ004801		
VERSION	AJ004801.1	GI:2653291	
KEYWORDS	complete genome.		
SOURCE	Bovine herpesvirus type 1.1		
ORGANISM	Bovine herpesvirus type 1.1		
REFERENCE	1 (bases 9702 to 103033)		
AUTHORS	Wirth,U.V., Fraefel,C., Vogt,B., Vlcek,C., Paces,V. and Schwytzer,M.		
TITLE	Immediate-early RNA 2.9 and early RNA 2.6 of bovine herpesvirus 1 are 3' coterminal and encode a putative zinc finger transactivator protein		
JOURNAL	J. Virol. 66 (5), 2763-2772 (1992)		
MEDLINE	94219360		
PUBMED	1313901		
REFERENCE	2 (bases 103034 to 111027; 127191 to 135300)		
AUTHORS	Schwytzer,M., Vlcek,C., Menekse,O., Fraefel,C. and Paces,V.		
TITLE	Promoter, spliced leader, and coding sequence for BICP4, the largest of the immediate-early proteins of bovine herpesvirus 1		
JOURNAL	Virology 197 (1), 349-357 (1993)		
MEDLINE	94025583		
PUBMED	8212570		
REFERENCE	3 (bases 111028 to 114234; 123984 to 127190)		
AUTHORS	Schwytzer,M., Wirth,U.V., Vogt,B. and Fraefel,C.		
TITLE	BICP22 of bovine herpesvirus 1 is encoded by a spliced 1.7 kb RNA which exhibits immediate early and late transcription kinetics		
JOURNAL	J. Gen. Virol. 75 (Pt 7), 1703-1711 (1994)		
MEDLINE	94292919		
PUBMED	8021599		
REFERENCE	4 (bases 66900 to 96900)		
AUTHORS	Vlcek,C., Benes,V., Lu,Z., Kutish,G.F., Paces,V., Rock,D., Letchworth,G.J. and Schwytzer,M.		
TITLE	Nucleotide sequence analysis of a 30-kb region of the bovine herpesvirus 1 genome which exhibits a colinear gene arrangement with the UL21 to UL4 genes of herpes simplex virus		
JOURNAL	Virology 210 (1), 100-108 (1995)		
MEDLINE	95313343		
PUBMED	7793062		
REFERENCE	5 (bases 1 to 31444)		
AUTHORS	Schwytzer,M., Styger,D., Vogt,B., Lowery,D.E., Simard,C., Laboisserie,S., Misra,V., Vlcek,C. and Paces,V.		
TITLE	Gene contents in a 31-kb segment at the left genome end of bovine herpesvirus-1		
JOURNAL	Vet. Microbiol. 53 (1-2), 67-77 (1996)		
MEDLINE	97164286		
PUBMED	9010999		
REFERENCE	6 (bases 1 to 135301)		
AUTHORS	Schwytzer,M., Paces,V., Letchworth,G.J., Misra,V., Bukh,H.J., Lowery,D.E., Simard,C., Bello,L.J., Thiry,E. and Vlcek,C.		
TITLE	Complete DNA sequence of bovine herpesvirus 1		
JOURNAL	Unpublished		
REFERENCE	7 (bases 30801 to 67800)		
AUTHORS	Schwytzer,M., Vlcek,C., Lowery,D.E., Bello,L.J., Meyer,G. and Misra,V.		
TITLE	Gene contents in a 37-kb segment centered in the UL part of the		
	bovine herpesvirus 1 genome: the last gap		
	Unpublished		
	Accession# Z78205		
	8 (bases 96901 to 99695)		
	Letchworth,G.J. and Kutish,G.F.		
	DNA sequence of the BHV-1 UL1 to UL3.5 genes		
	Unpublished		
	9 (bases 114235 to 122983)		
	Goltz,M., Bukh,H.J., Broll,H., Lewin,M., Mankertz,A., Boerner,B., Borchers,K. and Weigelt,W.		
	Nucleotide sequence of the HindIII O and K fragments located in the US region of the bovine herpesvirus 1 genome		
	Unpublished		
	Accession# Z98199		
	10 (bases 121402 to 123983)		
	Schwytzer,M.		
	Glycoprotein E and US9 genes of BHV1		
	Unpublished		
	11 (bases 1 to 135301)		
	Schwytzer,M.		
	Direct Submission		
	Submitted (08-SEP-1997) Institute of Virology, Faculty of Veterinary Medicine, University of Zurich, Winterthurerstrasse 266A, Zurich CH-8057, Switzerland		
	Location/Qualifiers		
	order(1..4014,102918..111027,127191..135300)		
	/organism="Bovine herpesvirus type 1.2 (strain K22)"		
	/vifion		
	/mol_type="genomic DNA"		
	/strain="K22"		
	/db_xref="taxon:31519"		
	order(4015..9664,15010..22094,26636..102917,114235..123983,135301)		
	/organism="Bovine herpesvirus type 1.1 (strain Cooper)"		
	/vifion		
	/mol_type="genomic DNA"		
	/strain="Cooper"		
	/db_xref="taxon:10323"		
	9665..11962		
	/organism="Bovine herpesvirus type 1.1 (strain P8-2)"		
	/vifion		
	/mol_type="genomic DNA"		
	/db_xref="taxon:10324"		
	order(11963..15009,22095..26635)		
	/organism="Bovine herpesvirus type 1.1 (strain 34)"		
	/vifion		
	/mol_type="genomic DNA"		
	/strain="34"		
	/db_xref="taxon:31517"		
	order(111028..114234,123984..127190)		
	/organism="Bovine herpesvirus type 1.1 (strain Jura)"		
	/vifion		
	/mol_type="genomic DNA"		
	/strain="Jura"		
	/db_xref="taxon:31518"		
	<1..472		
	/note="3'part of intron spliced out from transcript over covalently joined genome ends, leading to immediate-early RNA 1.5 and circ protein"		
	235..430		
	/rpt_type=DIRECT		
	/rpt_unit="235..248"		
	466..1280		
	/product="LRL1 (circ)"		
	/note="transcribed from alternative late promoter"		
	/evidence=experimental		
	486..1229		
	/note="immediate-early and late myristylated protein (vifion)"		
	/codon_start=1		
	/product="circ protein"		
	/protein_id="CAA06078.1"		

```

/db_xref="GI:2653292"
/db_xref="SPTREMBL:Q01342"
/translation="MGARASAPAAPPPHAAVLIDLALSGGTIDLPGGDEAVFVSCPTT
RPVYHMRGRGTHHTVPHFVGRAYAILPCRKFWLYLMRGSAVYGVPTTGLRLADS
LHDFLTATAGQOORDLHCLDVTVLDAQMDPVTFTPEILILEADPAFPPPPSARARS
TLRASRRPARFCFQHLLAAGSILDLCSPEQNAAPCSLLPACDSDAACPCDAGE
TARDCTADAAAPSPGALSRYSSVRVFF"
gene
1259..1264
/feature="circular"
1259..1264
/feature="circular"
/feature="circular"
complement(1649..1654)
/feature="UL52"
/feature="synonyms: UL53, UL54"
complement(1649..1654)
/feature="UL52"
complement(1658..2860)
/feature="UL54"
complement(1658..2860)
/feature="UL54"
/feature="early BICP27 affecting mRNA processing (cell
nucleus)"
/codon_start=1
/product="UL54 protein"
/protein_id="CAA06079.1"
/db_xref="GI:2653293"
/db_xref="GOA:Q01354"
/db_xref="SPTREMBL:Q01354"
/translation="MADPEIATLSTASESDSLFGSDREDDDEAPSLAPALRSVVGQ
VRKKLGGDEDEMPAPPGEGASGGPAEAPPARRARVRPRRRPRRRPRRPPAGE
QRSGPAAKEALATSHGGGGAARSIGSLRLARSLAEAAQORATATVAVFAG
RLDMLVQNGGFRAGVSWAALVDFGAQFVPEGRRTVTWETLMFHRDLYRMFEVR
PAAOARALRLCYLRSANLVDALASADECLTWCKFTATKNLRKTDPIVATAGAVL
ENLRKLAPRLCYLREGRPSLEELCAARLSLATCPASYMFMLARLSRAVRSAGE
CVPLLEVVDGAPFEETPGTCVAGLIDALDTHQACDSMTCKLVANFTLVPVIMHGK
YFYCNEIF"
complement(3040..4038)
/feature="UL53"
complement(3040..4038)
/feature="UL53"
/codon_start=1
/product="glycoprotein K"
/protein_id="CAA06080.1"
/db_xref="GI:2653294"
/db_xref="GOA:Q65816"
/db_xref="SPTREMBL:Q65816"
/translation="MLLGGRTVNLALALHTAHLALVALAARQACVATARN
GSURLEURSGAVTVGGANNATLADAPCRHAVVQHPGLDGDGAEALHGRVRAVAG
ARDCRAYLWCAQAGGLLAWLYVAVFYLRQERMFGLCRNDADFLSPGYTTLNAAA
ALAAVGHGPTTKLARLMCELSARRALAVDFRLDPLGCARWRPRAALPLIAEGFARLG
ARTAAAGSVGITHPCAAVPLYLKIWAHVVALPAGLEVLVLLYRPRRRGGTCAGDG
GDGSGSIRKVCVNCSTLGLLVKALYLAIVGGVIALHYEHLRLRLGCAQT"
complement(4013..7237)
/feature="UL52"
complement(4013..7237)
/feature="UL52"
/codon_start=1
/product="component of DNA helicase/primase complex"
/protein_id="CAA06081.1"
/db_xref="GI:2653295"
/db_xref="GOA:Q65817"
/db_xref="SPTREMBL:Q65817"
/translation="MDGPAALEDALQILYATDCAITGFSUMLLTGDEPGRGAVVVS
NWPGLAACLGSSPNDAADARPLVAFYLLGGPPGSSAAATRLARLRLVHLQOR
AGRARARRTPRQGRTRARAARPRGRHLCAARGHDHALLVATSQLAPRTQRTADD
ARVVEGTTAPSRARPAGRGSLASLIHHEKTVAAVYRRLYSNSGATPFWFLSKFG
PGSKTLVATFTFVQARAGADATDQVDRCLATYVAAPAPNSGLAFPLDVSAFA
ALAAFCRSYAGVAAGAPVVAARIADLAVERLREYIDHRRSLKVADEPFA
VYVLAYFEGNRRQITEHLRAVTAPEPPDGGGALAPRLAGVSLRERAVDAFPHVRA

```

Alignment Scores:
Pred. No.:
Score:

0.000153 Length: 135301
662.00 Matches: 656

Percent Similarity: 32.39% Conservative: 305
Best Local Similarity: 22.11% Mismatches: 1148
Query Match: 5.01% Indels: 870
DB: 14 Gaps: 138

US-09-522-753-5 (1-2517) x BHVICGEN (1-135301)

Qy 31 AlaArgThrHisThrAspValGlyLeuLeuGlu---TyrGlnHisHisSerArgAspTyr 49
Db 36388 GCGCGGCACGAGCTTAAAGTAGGCTTCGTAGAGCGGCATCCGACGACACGCA-GATCCC 36330
Qy 50 AlaSerHisLeuSerProGly-----SerIleIleGlnProGlnArgArg 65
Db 36329 GGAAGGCGCGTGCCTCGCGCGCCGCGCGGAGCTCGGCAACCTCGCGTCGAAGG 36270
Qy 66 ProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArg 85
Db 36269 CCG-----CCAGGAGGGCGCGCAGCCCTCTAGCGCGCTCGC 36231
Qy 86 ProGluSerHisSerTyrLeuProGluLeuGly-----LysSerGluMetGluPhe 102
Db 36230 CGCGCACCGCGCAAGCGTTGCTGACGCGCTGCAGCGCGCGCTCGCGCGCGCGCGG 36171
Qy 103 IleGluSerLysArgProArgLeuLeuLeuLeuProAspProLeuLeuArgProSerPro 122
Db 36170 CTTCTGTGAAGGCGCGCGCGCGCTCGCGCGCTCGC---CCAGCGCGCGCTCCAGCGCGC 36114
Qy 123 LeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThr 142
Db 36113 CTGTGAACG-----CGCGCGCGCGCGCGCG-----CGTAGTGTCTGT 36078
Qy 143 GlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeuGluLeu 162
Db 36077 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGG 36018
Qy 163 ValPro-----ProArgLeuSerLysGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 180
Db 36017 CGCGCTGTGTCTCTCGCGCGCGCGCG-----35991
Qy 181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeu 200
Db 35991 -----35991
Qy 201 GluGluGluAlaAlaLysProProGlu-----ProGluLysProValSerProPro 217
Db 35990 ---CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 35934
Qy 218 ProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLys 237
Db 35933 CCGAC-TCCAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 35888
Qy 238 AlaGluAlaAlaHisArgIleLeuGluGlyLeu---GlyProGlnValGluLeuProLeu 256
Db 35887 GCGCAGCGCGCTCGATCGCTCCGCAAGGGCTCATGGCGCGCGCGCGCGCGCGCGCGCGCG 35840
Qy 257 TyrAsnGlnProSerAspThrArg-----GlnTyrHisGluAsnIleLysIle 272
Db 35839 -----GTCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 35792
Qy 273 AsnGluAlaMetArgLysLysLeuIleLeuTyrPhelysArgArgAsnHisAlaArgLys 292
Db 35791 CTGACGACCTCCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 35735
Qy 293 GlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuLeuLysLys 312
Db 35734 CAG-----CAGCGC-----CGTTGCGTCAACGACGTCGCGCGCGCGCGCGCGCG 35699
Qy 313 ValGluArgIleGluAsnAsnProArgArgAlaLysGluSerLysValArgGluTyr 332
Db 35698 GCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 35666
Qy 333 TyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSer 352
Db -----352

Db 35665 -----CGCGCGCGCGTGCATCGCAGCAGGTTCTTGAG---- 35633
Qy 353 ArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluVal 372
Db 35632 ---GTTGGCAGCGCGCGCCCTGTCGCGGTGCGCGCGCGCTCGGCAGCACGACCTC 35576
Qy 373 SerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGlu----- 386
Db 35575 GCGCAG-----GATCGCTCGCGCGCGCGGTGGCTCGCGCGCGCGCTTTCCGCG 35522
Qy 387 -----LysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGln 404
Db 35521 CTCTAGCGCTCTCGCGCGCGCGCGCTCGCGCGCGCGT-----GCT 35477
Qy 405 GlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLys 424
Db 35476 CAGCATCTGGCGCTCTAGAGCGCGCGCTCGCGCGCGCGTAAAGCGCTCCAG 35420
Qy 425 AspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPhe 444
Db 35419 CTGGCGCAGCGCTCGCATCAGAGCTCGCTCGCGCAGCTCCAGCGCGCAGCGCTTC 35360
Qy 445 MetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAla 464
Db 35359 CAGCGCGCGCG-----CTCGCGCGCGCTGCGCAGTACGTACCTTCTTAGCGCGCT 35306
Qy 465 Glu----- 465
Db 35305 GATCTCGCGCGCGCGCGCGCGCTCGCAGCAGCTCGCGCGCGCGCGCGCGCGCG 35246
Qy 466 -----CysValLeuTyrTyrTyrLeuThrLys 474
Db 35245 CGCGCGCTCTCTCGAAGCGCGGTGCCCGCGGTAGAGCGCACGTACTTGGCGAG 35186
Qy 475 LysAsnGlu-----AsnTyrLysSerLeuValArgArgSer 486
Db 35185 CCAGGCGCAGCGCGCGCGCTCCAGGTACGTTCAGCGCTCGTAGTAGCGAGCTC 35126
Qy 487 TyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 506
Db 35125 GCGCGCTCCGAGCGCGGACGTGACAGCGCGCGGTGCGCATGTCATCAACCGGTGAG 35066
Qy 507 -----GlnGlnGlnGlnProMetProArgSerSerGlnGluLysAsp 521
Db 35065 CTGSCCCACGCTCGCGCGCGGAGAGCGCGCAGCGCGCGCGCGCGCGCGCGGAAACGC 35006
Qy 522 GluLysGluLysGluLysGluAlaGluLysGluGluLysProGluValGluAsnAsp 541
Db 35005 GTCCACACGCGAT---GCGCGCAGCTCGCGCGCGCGCGCGCGGACATC-----GAT 34955
Qy 542 LysGluAspLeuLysGluLysThrAspThrSerGlyGluAspAsnAspGluLys 561
Db 34954 CTTCGCGTTCTCGGCGCGGTGGGTGAACGCGAGCAGCGCTCGATGCTCGCGTCCG 34895
Qy 562 -Glu---AlaValAserLysGlyArgLysThrAlaAsnSerGlnGlyArg----- 577
Db 34894 CCGCTGCGGTCTCGCTCCAGCGCGCGTGGCGTGGCGCGCGCGCGCGCGCGGTGCTA 34835
Qy 578 ----ArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGlu---GluAl 595
Db 34834 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCGCGCGCGGAGTT 34775
Qy 595 alIeThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerArgTr 615
Db 34774 GCGCTCAGCGCGTCCAGCGCGCGCGCGCAAGTCTCGCGCGCGCGCTCTCTCGCGCTG 34715
Qy 615 pThrGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSe 635
Db 34714 CGCG-----GCCGCCCTCGCGCGCGCTCGCGCTCGCGCGCGCGCGCGCGCGACTTC 34661
Qy 635 rAlaIleAlaArgMetValGlySerLysThrVal-----SerGlnCysLysAsnPheTy 653
Db 34660 CTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 34601

Qy 653 rPheAsnTyrIlyLysArgGlnAsnLeuAspGluLeuGlnGlnHisLysLeuLysMe 673
Db 34600 CTGTGCGCCCAAGCGCGCGCGCGCGCATCGCGCGCGCGCTCGCGCTCGCTCGCT 34541
Qy 673 tGluLys-----GluArgAsnAl 679
Db 34540 CGCATCTCGCGCGCGTCCGCATCTCGCGCGCGCGCGCACCGCGCGCTCGTCGCGCG 34481
Qy 679 aArgArgLysLysLysAlaProAlaAlaAlaSerGluGluAlaAla----- 695
Db 34480 CCGAAGTACTTCTCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCTTGT 34421
Qy 696 -----PheProProValValGluAspGluMetGluAlaSerGlyValse 711
Db 34420 TTTCGCGCGCATCAGCGCGCGCTCATGGCGTTTTAGTGCCTCGCGCGCAGCAAG 34361
Qy 711 rGlyAsnGluGluMetValGluGluAlaGlu----- 722
Db 34360 CGCGCGCGAAGCTCGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGTG 34301
Qy 723 -----AlaLeuHisAlaSerGlyAsnGluValProArgL 734
Db 34300 CGCTCGCGCGCGCGCGCTACTGGCGCTGCGAGCGGTCCAAAACGCAAACTCGGCGG 34241
Qy 734 yGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerPr 754
Db 34240 GAGTGGCGCGCGCGCGCGCGCGCTCGCGCGCTCGCGCGCGCGCTCGCGCG-- 34183
Qy 754 oHisThrGluAlaAlaLysAspThrGlyGlnAsnGly-Pro-----L 768
Db 34182 -----GGGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 34133
Qy 768 yAspPro-----AlaThrLeuGlyAlaAspGlyProProProG 781
Db 34132 GCGCGCGCGCGCTCGCGTGGTCCCTCGAAAACGCGCGCGCGCGCGCGCGCGCGCTG 34073
Qy 781 lYPro-----ProThrProProArgArgThrSerArgAla---ProIleGluProT 797
Db 34072 GCGCGCGCGGTGCCCGACCGCGCGAAGGTGGCGCGCGCGCGCGCGCGCGCGCGCG 34013
Qy 797 hrProAlaSerGluAlaThrGlyAlaPro-----ThrProP 809
Db 34012 CGCGCGCGCGCTCGGGTTCTCCAGCGCGCTCGCGGTACCGCGCGCGCGCGCGCGCG 33953
Qy 809 roProAlaProProSerProSerAlaProProProValProLysGluLysGluG 829
Db 33952 GTCTCTGACCGCGCGCGCGCGCGCGCGCGCGCGCT-----CCTGC-AAGAAGCGCGCG 33900
Qy 829 luGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProProAlaAlaG 849
Db 33899 AGTCTCTGCGCTCTCGCGCGCGCGCTCTGCGCACAGCGCTCCAGCGCGCGCG-- 33847
Qy 849 luGluLeuAlaValAspThrGlyLysAlaGluProValLysSerGluCysThr--- 867
Db 33846 -----GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 33810
Qy 868 -----GluGluAlaGluGluGlyProAlaLysGlyLysA 879
Db 33809 GCNAACAGCGCGTCCAGCGCTGTCAGCTCTCGCGCTCGCTCCCGCGCGCGCGCG 33750
Qy 879 spAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysGluGlyG 899
Db 33749 GCGCGCGCGCGCGCGCGCTGCGCTCTAGCGCGCGCGCGCGCGCGCGCGCGCGCG 33690
Qy 899 lySerGlyArgAlaThrAla-----LysSerSerG 910
Db 33689 GCGCTCGCGGTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 33630
Qy 910 lyAlaProGlnAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaG 930
Db 33629 CGGCACCGCGCGCGCGCGCTCGCGCGG-----TCTGCGCGCTCTCGCGCGCGCGCG 33576

Db 31676 AGTGCACTAGCGCGCCGCCACCGGCACTCGGGCTGGCCACACGCGCGCAGAGCTAGGT 31617
Qy 1479 -----SerProGlyArgThrPhePro--- 1485
Db 31616 ACAGGTAGAGGCGCTCGGGGGCATCGCGCGAACAACGCTGGGCG-ACGTCCCGCTC 31558
Qy 1486 -----ProValHisProLeu 1490
Db 31557 CCATCGGGTCGAGAGTAGACCGCGTCGCGCGCGGAAGACACCGAGCCCATGGCG 31498
Qy 1491 AspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSer 1510
Db 31497 CCACGATCACCAGCGCCAGCGAGTCCGCGCGGTGCGCAGACGCGTCCACGAATCG 31438
Qy 1511 ArgProGlyThr-----AlaSerSerSerGlyGly----- 1520
Db 31437 CGCGCTGGGACTGGGTGCTCAGGGCGCGCTCGCGCGGGCGGTGAGAAAGCGCGCACTCC 31378
Qy 1521 -----SerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysPro 1536
Db 31377 CCATACACGCGCAAAACACGACGAGAGCCGCTCCCGCGCCCGCTGGTGGTGATG 31318
Qy 1537 ArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArg 1556
Db 31317 CGGTTTGGCAGGTGCACGATCGAG-----CACATTCCGGG 31282
Qy 1557 GlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSer 1576
Db 31281 CGCGCCCGCTTTGCGCGCCACGCGCGCCCTCAGAGAGACACCGTCGATGGCGTCG 31222
Qy 1577 SerIleAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerPro 1596
Db 31221 GCSCCAGCGCG-----GCGTCAGCGCGCCCGCGAAGCCACGCGCAGA 31177
Qy 1597 HisSerThrValProGluHisHis-----ProHisProIleSerProTyrGlu 1612
Db 31176 AAGCAGACGAGGACCGAAGGATGCGCAGCGCTGCGCGGCC-----AGCGCGCGTCG 31120
Qy 1613 HisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPhe 1632
Db 31119 TACTGTTGCGAAACCCACAGCGGGGACCGCGACGTGCGCGCGCGTTCGCGCGCGCG 31060
Qy 1633 AspProThrSerIleProArgGlyIleProLeu----- 1643
Db 31059 GCGCGCGCAAG-----CGGACATGGTCCGCTTCCGCTCCACAGGGGTACTAGGGGC 31003
Qy 1644 AspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThr-TyrProHis 1663
Db 31002 ACCGCGCGCGCAGCGCGCTTTATCGAAGGCCTCCGCGCGCACCGCGCGCGCTCA 30943
Qy 1663 sLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgG1 1683
Db 30942 CCGGGTGGCGCGCGGCGATGCGCGCAACCGCATCGACTTCGGCG-----AGCAGGTC 30889
Qy 1683 nThrIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMe 1703
Db 30888 TTCGTGTAAGC-----CACGCGCCCGCGCGCGTGT 30856
Qy 1703 tAlaGlnArgAlaAspMetLeu-ArgGlyLeuSerProArgGluSerSerLeuAlaLeuA 1723
Db 30855 TGCGCGCGCGCAGACGCGCGCGCT-----CGCGCCCGCGCAGACGCGCGGC----- 30809
Qy 1723 snTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGln----- 1736
Db 30808 -----GCTGACGCGCGCGCGGTCTACCGCGCCAGTCGCGCGCACGTCGCGCGGGGTC 30754
Qy 1737 -----ValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaM 1755
Db 30753 CAGGAGCGCGCGTCCCTCGTCTC-----CCGCAACCTCGGGTAGCCCT----- 30707
Qy 1755 etAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSers 1775
Db 30706 -----CCACAGCGCAGCAGCGCC-TCTTTCGCGCGCGCGCGCT 30668

Qy 1775 exProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrSerSerSerg 1795
Db 30667 TTGCGCGCGTGGCGCGCGCCCGGGCGGTACGCGGGGCTCTCTCGGACCTCGTCCC 30608
Qy 1795 luArgGluArgAspArgAspArgGluArgAspArgGluArgGluLysSerIleL 1815
Db 30607 AGACGCGCGAGCGCGCGCGAGCTCGCGACGTCGCGC-----TCGCGCGCAC 30557
Qy 1815 euThrSerThrThrValGluHisAlaProIleThrArgProGlyThrGluGlnSerS 1835
Db 30556 TGTGCTGTAGCGCGCGCGCGAGCGCGCCCATCG--CGTCCAGCGCTGCAGAAAGCG 30499
Qy 1835 exGlySerSerGlySerSerGlyGlyGlySerSerSerArgProAlaSerHis 1855
Db 30498 GCCC-ACCTCGTGGAGCGCGCGCGCTCGCGCGCGCAGAGCTTGGCGCGCGCAGCG 30440
Qy 1855 exHisAla-HisGlnHisSerProIle-----SerProArgThrGlnAspAlaLeu 1871
Db 30439 CCAGCGCGCTCTCGCGCGCGCAGCGCGAGTGGCGCGCGCGCGCGCTGTCT--- 30383
Qy 1872 GlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGlu 1891
Db 30382 ---CGTGTGTTCAGCGCGCGCGCGCGCGCGCGCGCTCGCGCGCTTCACT--- 30335
Qy 1892 ProSerLysProThrValLeuArgSerThrSerThrSerProValArgProAlaAla 1911
Db 30334 -----CCTGCGCGCTT-----GTGCTGTCTCGCGCGCGCGCGCTGTCT 30296
Qy 1912 ThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThr 1931
Db 30295 ACG---CCGCCAGCGACACGCGCACCTCGCGCGCGCG----- 30263
Qy 1932 LeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgPro 1951
Db 30262 ---CGCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGT-----CGGCAGGTTCGCGCG 30212
Qy 1952 ArgAlaAsp-----ThrGlyHisAlaPheLeu 1960
Db 30211 AGCGCCACCCAGCGCGCGCGCGCGCGCGCTATATCGCGCAGCATCTGTCT 30152
Qy 1961 AlaLysProProAlaArgSerGlyLeuGlu-----ProAlaSerSerProSerLys 1977
Db 30151 GCGCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30092
Qy 1978 GlySerGluProArgProLeu-----ValProProValSerGlyHisAla 1992
Db 30091 CACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30032
Qy 1993 ThrIleAla-----ArgThrProAlaLysAsnLeuAlaProHisHisAlaSer 2008
Db 30031 CCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29978
Qy 2009 ProAspProProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSer 2028
Db 29977 -----CGCCCGCGCGCGCGCGCG----- 29960
Qy 2029 LysProPheSerIleGlnGluLeuLeuArgSerLeuGlyTyrHisGlySerSerTyr 2048
Db 29959 -----CGGCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29909
Qy 2049 SerProGluGlyValGluProValSerProVal---SerSerProSerLeuThrHisAsp 2067
Db 29908 CTACCGTGTGCGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29861
Qy 2068 LysGlyLeuProLysHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArg 2087
Db 29860 -----TCGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29834
Qy 2088 Pro-----LysGlnProGlyProValLysLeuGlyGlyGlyGlyGlyGlyGlyGly 2102
Db 29833 CCACAAACGCGCAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29774

```

Qy 2103 LeuProHisLeuArgProLeuPro-----GluSerGlnProSer 2115
Db 29773 TGGCCA-----CGCTCCCGCGCGAGTCTCGCGCCCAAGCGATCAGCGCCATCA 29723
Qy 2116 SerSerPro---LeuLeuGln-ThrAlaProGlyValIysGlyHisGlnArgValTh 2134
Db 29722 GCTCGGCACGCCGTTTGGCGTGAAGGCGACGCGTCCGAGCCACGACGACGTC----- 29668
Qy 2134 rLeuAlaGlnHis---IleSerGluVal-----IleThrGlnAspTyrTh 2148
Db 29667 ----CGACAGCAGCGCGCGCGAGTGGCGGCAAGCGCGTCTCGCAAGACGCGAG 29612
Qy 2148 rArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProG1 2168
Db 29611 CCGGACAGCGGCTCGCGCGG-GGAACTCGAGCGCGCGCGC----- 29570
Qy 2168 yAlaSerCysProValLeuAspLeuArgProProSerAspLeuTyrLeuProProPr 2188
Db 29569 -CGCGGGCGAGCTCTGCGCGCGAAGCGCGCGCGCGCGCGAGAGCGTCCGCGCC 29511
Qy 2188 oAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProG1 2208
Db 29510 GCGGGCGCGCTTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29451
Qy 2208 uProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProPr 2228
Db 29450 AACCAGG-----TTAGAGGCGCTCGC-----CGCGCGCAGCTCCGCGCGCC 29409
Qy 2228 oGluGly-----MetThrGluProGlyHisSerArgSerAlaVal----- 2241
Db 29408 GCGCGCTCGCGCTGTTTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCGCTGCTCC 29349
Qy 2242 -----TyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMe 2257
Db 29348 GCGCGCGCGCGCTTCGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29289
Qy 2257 tGlySerLysSerProGlyAsnThrSerGlnPro-ProAlaPhePheSerLysLeuThrG 2277
Db 29288 CGCGGGCCATGCAGTGCAGTCTGTCTCCAGCGCGCGCGCGCGCGCGCGCGCGCG 29229
Qy 2277 luSerAsnSerAlaMetValLysSerLysLysGlnGluLeuAsnLysLeuAsnThrH 2297
Db 29228 GGCAGCAGCTCCGCGTAGTACGGGTGCGAGTTTCGATAGTGGCGGTGTAGTGGCGAGG 29169
Qy 2297 isAsnArgAsnGluPro-----GluTyr-AsnIleSerGlnProGlyThrGlu 2312
Db 29168 TAGTCCCACTGGCGCCGTCAGCGCTCAAGAGTTTAAACAGCAGCGCGCGCGCGCGCG 29109
Qy 2313 IlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVal 2332
Db 29108 ACGCAGCCACCGCGCTGCAGCAGCTCTGCCGCGC-----CGCGGCTCTGGCGCG 29058
Qy 2333 GlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGly 2352
Db 29057 GATAGCTGCCGCGCAGC----- 29040
Qy 2353 LysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsn 2372
Db 29039 -----CGCGCTCTAGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 28995
Qy 2373 AlaSerAlaSerLeuPro-----AlaAlaMetProIleThrAlaAlaAspGlyArg 2389
Db 28994 GCACGCTCTCCAGCGTTCGTCGGCGTTCGCGCTCCGTCGCGCGCGCTCCCAACCGCG 28935
Qy 2390 SerAspHisThrLeuThrSerProGlyGlyGlyLysLysValSerGlyArgPro 2409
Db 28934 GCGCGGCTCTGCAACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCT 28875
Qy 2410 SerSerArgLysAlaLysSerProAlaProGly---LeuAlaSerGlyAspArgPro 2428
Db 28874 AGCAGCGC-----ACCACCGCGCGCTCGCTGAAACGCGCGCGCGCGCGCGCG 28824
Qy 2429 SerValSerSerValHisSerGluGlyAspCysAsnArgArgThr-----Pro 2444

```

```

Db 28923 AACGTGTCTGTCGGCGAGAGCGCGCTCGCGCAGCGCGCGCGCGCGCGCTGCCCGCG 28764
Qy 2445 LeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyr 2464
Db 28763 GCGCGCGCGCGCACAAAGGGTCCCGTAGCCAGCGCGCGCGCGCGCGCGCGCGCGCG 28704
Qy 2465 AsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeu 2484
Db 28703 GACGAG-----CGGCAAAAGCGTTAGCCTCA----- 28677
Qy 2485 ProAlaGlySer-GlyProLeuAlaGlyProHisAlaTrpAspGluGluProLysPr 2504
Db 28676 CCGCGCGCGCGCTAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 28617
Qy 2504 o 2504
Db 28616 A 28616

RESULT 94
SFU78289/c
LOCUS
DEFINITION
SFU78289 43280 bp DNA linear BCT 13-AUG-1997
Streptomyces fradiae tyactone synthase, starter module and modules
1-7, (tylg) gene, complete cds.
ACCESSION
VERSION
U78289.1 GI:2317859
KEYWORDS
SOURCE
Streptomyces fradiae
ORGANISM
Streptomyces fradiae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
1 (bases 1 to 43280)
DeHoff, B.S., Sutton, K.L. and Rosteck, P.R. Jr.
Sequence of Streptomyces fradiae tyactone synthase gene tylg
Unpublished (1996)
REFERENCE
2 (bases 1 to 43280)
DeHoff, B.S., Sutton, K.L. and Rosteck, P.R. Jr.
Direct Submission
TITLE
Submitted (14-NOV-1996) Eli Lilly and Company, Lilly Corporate
Center, Indianapolis, IN 46285, USA
tylg.
FEATURES
source
1. 43280
/organism="Streptomyces fradiae"
/mol_type="genomic DNA"
/db_xref="taxon:1906"
1. 43280
/genes="tylg"
816..14234
/genes="tylg"
/note="ORF1; putative"
/codon_start=1
/transl_table=11
/product="tyactone synthase starter module and modules 1
& 2"
/protein_id="AAB66504.1"
/db_xref="GI:2317860"
/translation="MSSALLRAVQSCGYGDLMTSNTAAQNTQDEVDGPDSTHGGE
IAVWNSCLPLPGAAGVEEFWELLSRGMPTRQDDGTWRAALDHAGFDAGFGMNA
QAAATPOHRLAMELEAGELVPGDITGDTGVFAGVADYAVLTVRRSAVSAG
GYATGLHRLAANRLSHFLGRLPSLVDSQASLVAVOLACELRRRGETSLAVAG
GYNLILTESTTWMEWGALSPPGRCTHEDARANGYVRGEGGAVVLKPLDALADGD
RYCYVIGKANVNDGGASLTTPDREANGVLRQAYRRAGVSTGAVRYVLEHTGTGA
GDVEAALGALVLAGADSGRTPLAVGSKTVKTHLGAAGIVGLIKATLCVRKGE
VPSLNSFTNPDIPLDLRLVQTERVWNEEDRPRVAGSVSGFMGTGNHVLIAE
PAAAGSGAGGAGGAGCAGISAVSGVVPVNSGRVSVVREAGRLAEVVEAGVGIA
DVATWADRSRFGYRAVLARGAEALAGRLALAGAGDPDAGVTVGAVLDGCVVGAAP
GAGAGAGGAGGAGGCGVLYFCGCTQWVCGMAGILLGSSSVPRASMECARLSV
HVGNDLLEVVSGAGLERVDVQPTWAVNVSRLRWQAMGVDAVAVVGHSGQEIAA
TVAGALUSLEDAAVVALRAGLIGRYLAGMAAVPLPAGVEAGLAKWPGVEAAVN
TPASTVSGDRAAVAGVACQAEQVABRLIPVDYASHSHRVEDLKGELRRVLSGTRP
RSRPRVPCSTVAGEQPEPFDAGYWRNLNRNRFVSAVVGGLEGHRRFIEVSAHP
VLVHAIEQTAEAADRSHVATGTLRRQDDSPHRLLTSTAEAWAHGATLTWDPALPPGHL

```

TTLTPTYPFNHHYWLDTIDGGGDDATQEKESGQLTRELRLPSSQKQGLFLDLVCR
HTAVLGLDTAAEVPDLSFKQKIOSMTQVLEARNRLLTETGLALTTLVYDRPTPRA
QDAFLATELLDGSPSCSVLAPAKSFAEQEPIAVCMGCRFPGGVSGPEALWFLVVEG
LVAVSPFPGDGMVDGLVDPEFVAGKSTVREGGFHDAAEFDAEFFGLSPREAVAM
DPQORLLLETSWAIERAGIDPHSLHSRTGVYAGVYGRPLAEGMAEGSDGYLLT
GTSQSVVSGRVAIVTGLDPAVTVTACSSILVHLVAQALRGCECDMAAGGVTVM
VAGPFVFSRQGLAADRCRKAFAADGAGVAVLVERLSDARRLGHPLVAV
CCGSVNODCASNGLTAPSPSOBVIROALGNARLTADVDDVVEAHGTRGLRPIE
AQALLGTYGRDRGGRPVWLSIKSNIHQAAAGVAGVIMWLANRYGWLPTLHVD
EPRHVDMSAGGWWLITAEKWEFGVDPRRAVSAFGSGTNAHLILEAPDPAESA
TTPVRSEVSAVVDARSGVTVVSGRVSRRVVRRAAGLAELAEVVEAGGVDADAVT
MAGRSQFVGVAVLAGEALAGLALACAGDPDAGVTVGVVDDPETGGGGVGLVF
PGCGTQWVGAGLGLSGSEVFAASRECARALSVHVEWDLLEVVSGAGLERVDVQVP
VWAVYSLARVQWGMVDVAAVGHSGQEIATAVAGALSLEDAAVVALRAGLGR
YLAGRMAAVPLPAGEVEAGLAKWFGVEVAANGPASTVSGDRRAVAGVACQAE
GVQARLIPVDYASHSRHEDLKGELERVSIGIRSPRPVPCSTVAGEQPGFVDFAG
YFNRNLRNRYEFAVVGGLLEGHRRPIEVSAPHLVHAIEQTEAAEADRSVHATGT
RQDSPHRLTSTAFAWAHATITWDPALPGHLLTPTVFNHHYWLDTTPTTAT
TTQSPDAMRYRTWKALTESPVRSIGRCLLVAPPPTDGLDGLTTLVSER GAS
VARLEPIGARAEVAELLKPSMESAGEENTTVVSLGLVPSDVAVRTSIALQAVSD
IGVPAARVWALTRAVAVPGETPDQAGLQWGRFVAALELPDIINGGLIDLPEABL
TRTPETSQPTPERIPQTPNRRALLEAAVLAGRDEQDVAVRASGIVGRRVSRAAA
AGASQWSPQTVLITGCGAIGRRLARLAEAGERLVLTSRRGPEAPGAELAEELR
GHCEVHAAACDVAERDAAALVAVPNVAFHTAGLDDDAVIDTLSPESFVRAGL
VCCGELLHOLTADIKGLDAFVFSVTGTWGNAGQAGYAAANAALDALAEERRAAGL
ATSVAMGLNAGWAGAGESISSLRGLRAMDPDAVDALLGAMGRNDVCVTVDVM
ERPAATNALRPGGLFDTPVEAREALTAAGTTSATPDGAPELARRLSMLNETERLKL
VELVRETAARVHPNDTDAIGERPFKAGFDSLSLELRNLNAGTGLKLPATVIFD
HPSFTALRLLRLTGAGAPAADEPLPFVAVADDPVVIYVGMACRFPFGGATPEA
LWKLVTBEERDIAAFTDGMWDLSDVDDPFGVAGKTYVREGGFLHDAAEFADFEGL
SPREAVAMDPOQRLLETSWEATERRAGIDPHSLHSRTGVYGLTQHEVASRLHEAPE
EYGVLLTGKASVWSGRI SYTLTGLEGPSLS IDTACSSILVHLVAQALRGCECDMA
LAGVTVMAAPGLVFESRORGLAAGRCRKAFAADGAGVAVLVERLSDARR
LGHPLVAVCGSVAVNQDASNGLTAPSPSOBVIROALGNARLTADVDDVVEAHGTR
TRIGDPTAQLLGTGRDRAECPVWLSKSNIGHAQAAGVAGVIMWLANRYGWL
LPLRTLVDESRHVDMSAGGRLVLTTEAREWFGVDPRRAVSAFGSGTNAHLILEAP
EALAEALDAPAEPAEPDVTDTVEALEADPAEAGAKAPSGPEAPAGVGVVP
VVSGRSRVVRAAGLAEVVEAGVGLADVAVTWAGRSRFGVRAVVLARGAEALG
RLRALLAGDPDAGVTVGAVVDPETGGGGVAVLFPQCGTQWVGAGLGLSGSEVFA
SMRECARALSVHVEWDLLEVVSGAGLERVDVQPTVWAVVSLARYQWGMVDVAAV
VGHSGQEIATAVAGLALSLEDAAVVALRAGLIGRYLAGRGMAAVPLPAGEVEAGL
KWFGQVAAVNGPASTVSGDRRAVAGVAVCAEQVQARLI PVDYASHSRHVEDLKG
ELERSIGIRSPRPVPCSTVAGEQPGFVAGVFNRLNRVBFSAVVGGLLEBQ
HRRFIEVSAHPVLVHAIEQTEAADRSVHATITWDRQDSDPHRLTSTAFAWAHATL
TWDPALPPCHLTITPTYPFNHHYWLTVSPAGVGDAAAGFGMTWEDHPELRCGLPIA
DSGERVPAGLAGEHWDLTHAVSVGLLPGFATVFPALHAGAAATCCGLELSVEA
PLVLPAGVVRVQMRVSADESGRRVVAIHSAPAAVHSAAGGSDGSAVWVRHGEGL
VPDPEPTPDADWARWPPAGERVEPAELYERFALGYEYGEAFAGVAVVRQPDALL
AGALVRLTKAGDGVASLOADTSGAVALSISGALVMPPLARRKLVLLIGTDACERSLY
RVWQVRLLPAGPSPSWANLGPADRRLAGTFLGIDQDPDGGFTALYEVRLAKAAG
APREAVLPLVSGAGATPVSQVQTERCLTALQMDLDAEELVDPLILVLRGAVAV
PBEIGDLAGCAGVGLVRSARSHPGRFALVDTHGPHDDRTALPLALRAVLDGAGLS
LRAGATPTVLLRAGTPEORGFAPDPAGTVGTGTLGRLLARHLAEHGRVHLL
LGRGGDQADAPERLAEALAGLAEPCFAACDAADREALRAVLAEPADPLGTGTHA
AGVLDCTDALPTERTGTVMRKADALNHELTRTSPUSVAFSGSAGIILGRQCP
ANYAANTFLDALAQHRAGLPAVSLWGLMGATGTMGTLSGTDLRMRRSGLAPM
THOQGLAFDRAASAEDPLVPMRLDLAALVREAREHVGDAVPGFLGILLPARAV
RQAAVFRGAPAPAGEGTAERLAGEARLRELVRVRAEVSGLVSGPDVAP
GRPFKIDJFQSLTAVELNRIGATGLRLTALFPDRPTQSAVLAELAGLPRDGG
DTFAAQEGLEALAAAGALAEEDLRDLRLRLRLTELAALTLTQGRNPSAPAPASDL
DERLSDANDDLFAFIEQL"

942..3929
/gene="tylg"
/function="encodes starter module"

942..2156
/gene="tylg"
/function="encodes beta-ketoacyl ACP synthase KS-Q of starter module"

2571..3557
/gene="tylg"
/function="encodes acyltransferase domain of starter module"

misc_feature 3675..3929
/gene="tylg"
/function="encodes acyl carrier protein domain of starter module"

misc_feature 3993..8471
/gene="tylg"
/function="encodes module 1"

misc_feature 3993..5264
/gene="tylg"
/function="encodes beta-ketoacyl ACP synthase of module 1"

misc_feature 5631..6617
/gene="tylg"
/function="encodes acyltransferase domain of module 1"

misc_feature 7410..7949
/gene="tylg"
/function="encodes beta-ketoreductase domain of module 1"

misc_feature 8220..8471
/gene="tylg"
/function="encodes acyl carrier protein domain of module 1"

misc_feature 8541..13970
/gene="tylg"
/function="encodes module 2"

misc_feature 8541..9812
/gene="tylg"
/function="encodes beta-ketoacyl ACP synthase of module 2"

misc_feature 10260..11246
/gene="tylg"
/function="encodes acyltransferase domain of module 2"

misc_feature 11319..11876
/gene="tylg"
/function="encodes dehydratase domain of module 2"

misc_feature 12861..13415
/gene="tylg"
/function="encodes beta-ketoreductase domain of module 2"

misc_feature 13719..13970
/gene="tylg"
/function="encodes acyl carrier protein domain of module 2"

CDS 14351..19945
/gene="tylg"
/note="ORF2"
/codon_start=1
/product="tylactone synthase module 3"
/protein_id="AAB66505.1"
/db_xref="GI:2317861"
/translation="MTAEIVATKRLGALLEERAREPIAVVAMSCRYPGGVTTPEDLMR
LLADERDAVSGLPDRDGLDLDALYDPGGPSTYAREGGFLSHCAGFDAEFFGISPRE
ALAMDPOQRLLETSWEALERAGVTADRARGSRGVYAGVYDDYDARVLYGAGGPP
EDLEGVLNVSAGSIASGRVSYTFLGRPAVTNTACSSLSVSLHLVALRNGECDM
ALAGATVLTSTPVLVDFSRQGLADGCRKAFADADGTSWAEAGMLLQRLSDAR
REGRPVLAVIRGSVAVNQDASNGLTAPNQAQVIEDALRDAGVDGQDVAVEAGT
GTELGPIDRAGALLTGTARTAEERLPLWLSKNSIGHTQAAAGVAGVIMWLANRYG
RLPTLHVRPPTTRVDEWKGVRLLTPEVPWPGCEAGPRAGVSSFCASGTAAHVLE
SPVAGPAAAGREDTGGATVSGRPAALRAQALYDALTGTGTGTGCGACGAGP
GTAEVAGALAHARTAEHRAVILGGRNRELACGLRELEHEHPRVVTGTATPATER
TAFLFSGQSQRSGRGLVRRHPVPRARALDEVCALLEPHLHRLPLDFAEPGSPER
EPIDRTEFTOPALFALQALFLAEHGLRAELACGSGVETAAHAAGVITLIDPAAR
LVLAERLQWALPAGGMAALRATAEIIAPLERRAGELALAAVNGPSSVVVSGDEAA
VLEAQRIMQAEGRAREARLAVSHAFPRMDMDLTQFDRVARTLTFTPTLPTVLTGT
TPVTEETCTADHWQAREPVRFLDMRTLRADGIDTFLVGLPDGVLTSAMARDCAAD
RPGDITGAGDGETPDPLTLPLLRSSVPTGDAEHPGFERATATAYAHGVPLRLAP
APDAASLAVAAELPTVAFQTHYWDPAAPALPAGLDDAGHPLLSAALDLPGGRT
VWTKGUSAATLPWAADHSVHGRVTLFGTALLDLALHAAPRVGELTEAPLVLPEDEGT
RLRVLAEPDASRELSVHSGEDGSGWTHATAVLDTGTTTAGEFAGAPVPGD
AEPDLDAEYERFADAGIGYFAGRLRSARWDCDALADVRLTPAGDADAVAVDPA
ALLDAAHLTAAALGGAGMLPPTMNGVTLHARGAHAI RVLTPAGDADAVAVDPA
GRPVFTAASTLRLPTVTGOLTAEEAARAPLYRVVTGTPDTGTARLHTWAVAGPGDL
LPGETPHHPLASALADTGTAPERVLADLRYGCTATRELASOALALVQOADAABEA
EGRLVLTVRADVIGDGTDPAAAATWGLVRAAQSEHPCGFALLDITDSDSRQLPR
VAGTABQLARLDRLIAPSLRATLPAGARLPALDGTVLITGTGSLGAEAAHRLVTR
HGARRLLTSRSGPQAGAEELVAELAAHGAHADVAACDVADRAALRLDLRVPFAGH

Qy	1093	rgProValLeuProArgProThrIleSerAenProProProLeuIleSerSerAlaL	1113
Db	9234	--CGCCCCCATCCGCGAAGCCTTGCACCGCCATCGCGCGCAACCCCGCTCGCGCGA	9177
Qy	1113	ysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnL	1133
Db	9176	AAACTCCAGN-----ACAAATCCGGGTGCCGCCATCACCGTCAC	9138
Qy	1133	euhisValProTyzSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuP	1153
Db	9137	ACCACCGGCCAACGCCATGTGCACACTCGCCACCCCGCAACGCCTCGC-----	9091
Qy	1153	roLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuS	1173
Db	9090	-----CCGCGTTTGTACGGCGCACCGACGACGACCAACACCGCGGTGTC	9048
Qy	1173	erProArgGlyGlnAlaGlyPro-ProGlnSerLeuGlyValProThrAlaGlnGluAla	1192
Db	9047	GATGGAGGAGGAGGACCTCCAGCCCGGCGTGTACGAGATGGCGCGGAGACGACGT	8988
Qy	1193	SerValLeuArgGlyThrAlaLeuGlySerVal-----ProGly	1205
Db	8987	CGCGACTTGCCTGTAGCAGATAGCTTCGTACTCTCCGGGGCTCTGTGACGCCGGA	8928
Qy	1206	GlySerIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyr	1225
Db	8927	GGCGTACTCTGTATGGGTGACGCCGAC-----GTATAC	8895
Qy	1226	ArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArg	1245
Db	8894	CCCGGTGCG-----GCTGCCGTGCACGCGA	8871
Qy	1246	IleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLys	1265
Db	8870	GTG-----CGGGTCGATACCGCGCGCTCGATCCCTCCCTCCACGGA	8832
Qy	1266	Gly-----HisVal-IleTyrGluGlyLysLysGlyHisValle	1278
Db	8831	GGTCTCCAGCAACAGCCGCTGCTCGGATCCATCGCCACCGCTCACCGCGCGAAATCCC	8772
Qy	1278	uSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSer-SerS	1298
Db	8771	GAAGAACTCCGCGCTCGAACTCCGCGCGCTGTGGAGAAAACCCCTCCCGCACATGT	8712
Qy	1298	erGlyProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgV	1318
Db	8711	CTTCCCGCCACACCGCGTCCGGTCTGTAGACGGAATCCAGGTCCAGCCCCCGGTGGT	8652
Qy	1318	alGlyArgAlaIle-----SerSerAlaSerIle-----	1327
Db	8651	GGGCGCGCGCTATGATAGTCACGCTCTCGGTACCGACTTCCACAGCGCTCAGGGGT	8592
Qy	1328	-----GluGlyLeuMetGlyArgAlaIleProProGluArgHisSerP	1342
Db	8591	GCCCGCCCCACCGGGGAACGGCAGCGCATCGCGACGATCACCGGGTCTGTCTCGCGC	8532
Qy	1342	roHisHisLeuLysGluGlnHisIleArgGlySerIleThrGlnGlyIleProArgS	1362
Db	8531	CACGG-----	8527
Qy	1362	erTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArgGluG	1382
Db	8526	-----CGACGGCGAGTGGCGGTCTATCGGGCGCG-G	8497
Qy	1382	lyThrProProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaL	1402
Db	8496	GCGCGGGGGTCCGGCGCGGTACGCGGTGAGACGACGTCTGTGGCCAGCGCGGTCCGGC	8437
Qy	1402	euglyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaC	1422
Db	8436	TGGGGTGGTCCGAAGATG-----ACGGTGGCGGGTAGCT	8404

Qy	1422	lyArgSerIleHisGluIleProArg-----GluGluLeuArgHisThrProG	1433
Db	8403	TCAGCGCTGTCGGCGCATTCAGCGGTTTCGCGAGTCCAGGGAGGTTCGAAC	8344
Qy	1438	luLeuProLeuAlaProArgProLeuLysGluGlySerIle-----ThrG	1453
Db	8343	CGGCGGACTTGACGGCGGTTCCGGCCGATAGGCTCCGTTCGGATCGGCACAA	8284
Qy	1453	lnGlyThrProLeuLysTyzAspThrGlyAlaSerThrThrGlySerLysLysHis	1473
Db	8283	AGGCCCGCTCGGTACGG-----ACGAGTTCGACAGGTTCCGACGGCGTTCGG	8236
Qy	1473	al-----ArgSerLeuIleGlySerProGlyArgThrPheProP	1486
Db	8235	TCTCGTTACATGACAAACCGCGCCAGCTCGGGCGCGCTCCGGCGTC-----	8183
Qy	1486	roValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluAtrGAlaCys	1506
Db	8182	-----GCGAGCTGGTCCGCTCCGTACGGGCTCCCGGCCCTCCGGCAGG	8134
Qy	1506	luSerLeuLysSerArgProGly-----	1513
Db	8133	TGTCG---AACAGCGCCCGGACGATGCGTTCTGTCGGCGCGAAACGCTCCCA	8077
Qy	1514	-----ThrAlaAs	1516
Db	8076	CGAGTCGACGACAGTGCACACACGCTGTTCTCTGCCATGGCGCCAGGAGCGCT	8017
Qy	1516	erSerSerGlyGlySerIleAlaArgGly-----Alap	1527
Db	8016	CGGCGCGTTCGGGTTCATGGCCCGCAGCCCTCGCCGCGACAGACTCTCTCGC	7957
Qy	1527	roValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyzGluAs	1547
Db	7956	CGCGCCCATGCGCT-----CCCGCGCCCAAGAGCCCGCAGGCGCGCGG	7906
Qy	1547	lyAlaProPheAla-----	1551
Db	7905	GCAGTCGGCGGACGCGGACGCTCGCGCAGGCGCTCGAGCGCGCGCTTGGCG	7846
Qy	1552	-----GlyHisLeuProArgGlySerProValThrMetArgLup	1565
Db	7845	ACGCACCTCGCGCGGTTCGCCCATGTGCCGTGACGAGGAGAGACGAAGCG	7786
Qy	1565	roThrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAsp	1585
Db	7785	CCAACCTTTATGTCCGCACTCAGTTGGTGCAGCAGCTCCGCGCCGCAC-----	7736
Qy	1585	euThrSerThrProArgGluIleAlaLys-SerPro-----	1596
Db	7735	ACCTTCGCGCGCGGACGCTCTCGAAGCTCTCCGTGACAGCGTGTGCATCAC	7678
Qy	1596	-----	1596
Db	7677	CGTCCAGAAATCCCGCGGTGTGAAGACGCGTTTCGCGGATACGGGTGACGAG	7618
Qy	1597	-----HisSerThrValProGluHisHisProHisProIleSerProTy	1614
Db	7617	CGAGCGCATCAGCTCGGCCACGTGACAGCCCGGT---GCAGACCTCCGACCAT	7561
Qy	1615	Leu-----	1615
Db	7560	CTCGAGTTCCTCGGCGAGTTCGGCGCGCGCGCGCTCCGTCGCGACGCGTGT	7501
Qy	1616	-----ArgGlyValSerGlyValAspLeuTyzArgSerHisIleProLeu	1631
Db	7500	GGACCGAGCTTCGGGT-CCTCTCGCGCGCGCAGCGCTTCGGCGAGCGCGCT	7442
Qy	1632	PheAspPro-----ThrSerIleProArgGlyIleProLeuAspAlaAla	1648
Db	7441	CCCATGCCCGCGGTATCAGACCCCGTCCCGGACGGCTGCCAG---GAGG	7385
Qy	1649	TyrTyrlleuProArgHisLeuAlaProAsnProThrTyrlleuTyrlleuTy	1668

Db 7384 GCGCGGC-CCGGACAC-||| ||| ||| 7350
QY LeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAsp 1688
Db 7349 -CGAGGC-||| ||| ||| 7344
QY TyrIleThrSerGlnMethHisAsnThrAlaThrAlaMetAlaGlnArgAlaAsp 1708
Db 7343 -CGCACGCGACCTGGTCTCCGCGTCCGCGCCGCGGAGGAC 7302
QY 1709 -MetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAla 1725
Db 7301 GCGCGCAGCAAGCTCAAGGGCGGTGGT-TCGGAGTCTCGGCAGCCTCCGGGGTCT 7243
QY 1726 AlaGly-||| ||| ||| ProArgGlyIleIleAspLeuSer 1735
Db 7242 GTGGGGCTGTGAGGTCTCGGCGTCCGCGTCAGCTCCGCTGCTCCGGCAGATCGATCA 7183
QY 1736 GlnValProHisLeuPro--ValLeuValProProThrProGlyThrProAlaThrAla 1754
Db 7182 AGCCGCCCCAGATATCCGGAGTCAAGGCGCCCACTCGTCCGAAGCCCCCAACTGGG 7123
QY 1755 MetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSer 1774
Db 7122 CCCCCGCTCTCGGGGTCTCCCGGGAACACAGGCCACCGCCCTCCGCGTCAGCGCC 7063
QY 1775 SerProLeuSerProGly-||| ||| ||| 1780
Db 7062 AGACCTGGCGCGCGGACGCCGATGTCGAGAGCCGCTGGAGGAGCGTATCGACGTCC 7003
QY 1781 -GlyProThr 1783
Db 7002 TGACCGCGTCCGTGGAGGACACAGACCCGAGAGAGCAGAGCGGTGTCTCTCTCC 6943
QY 1784 HisLeuThr-||| ||| ||| 1786
Db 6942 CCCTGACTCCATGAGGGGTTTGACAGTTCCGCGACCTCGGCACCGCGCGCCGATGS 6883
QY 1787 -LysProThrThrThrSerSerSerGluArgGlu 1797
Db 6882 GCACCTCAAGCGCGGACCGAGGACCGCGCTCGGACACACCGTTGTCACTCGCTCGA 6823
QY 1798 ArgAspArg-||| ||| ||| AspArgGluArgAspArg 1806
Db 6822 GGAGCTCGCGTCTGGTGTGGGGGTGCAACAGAGGAGGAGCGACCGATGGAGTGGGC 6763
QY 1807 AspArgGluArgGlyLysSerIleLeuThrSer-||| ||| ||| ThrThrThrValGluHisAlaPro 1825
Db 6762 GGACGGGGAGGATTCGGTCAGGGCTTCCAGGTGACCGGTAGCGCCAGGCATCGG- 6706
QY 1826 IleTrp-||| ||| ||| 1827
Db 6705 -TGGGGCTCTGGGTGTCTCGCGGGGTGTGGGGTGTGTGTCTGAGCCAGTAGTGTGT 6649
QY 1827 -||| ||| ||| 1827
Db 6648 GGTGTTGAAGGGTAGGTGGGAGGGTGTGTGAGTGGCTTGGGGCAGCGCGGGTCCC 6589
QY 1828 -ArgProGlyThrGluGlnSerSerGlySerSerGly 1839
Db 6588 AGGTGAGGGTGGCGCGTGGGCCAGGCCTCGCGGTGGAGGTTCAGCAGCGGTGGCGGC 6529
QY 1840 SerSerGlyGlyGlyGlySer-||| ||| ||| SerSerArgPro 1851
Db 6528 TGTGCTCTGGCGCGCAGGGTCCCGGTGGCATGGACATCCCGTCCGCGCGCTCGCGCG 6469
QY 1852 AlaserHisSerHisAla-||| ||| ||| 1857
Db 6468 TCTGCTCGATCGATGACAGTACCGGTGGGCATGACCTCGATGAACCGAGCGGTGGC 6409
QY 1858 -HisGln-His-||| ||| ||| 1861

Db 6408 CTTCTCTCAACAAACCCAGCCAGCCGCGGAGAACTCAACCCGGTTCCGAGATTACGGA 6349
QY 1861 rProfile-||| ||| ||| 1863
Db 6348 ACCAATATCCCGCATCGAAACCCGGCTCGCCGGTCTCTCCGGCGAGCGGTGGAAACA 6289
QY 1864 -SerProArgThr-||| ||| ||| GlnAspAlaLeuGlnGlnArgProSe 1876
Db 6288 CCGGCACCGCGGACTCGGGGGCGGATACCGGACAGCAGCCGCTCCAACTCGCCCTTCA 6229
QY 1876 rValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSer- 1893
Db 6228 GGTCTCCACATGGCGGAGTGAGAGGGGTAGTCCACCGGTATCAACGAGCCTGCACAC 6169
QY 1894 -LysProThrValLeuArgSerThr-||| ||| ||| 1901
Db 6168 CTTCCGCTGACAGACGGCCACATACCGGCGCACCGCCCGCGATCCC-CGAAACACCC 6110
QY 1902 -SerThrSerSerProValArgPr 1909
Db 6109 GTGGAGCGCGGACCGTTGACCGCGCGGACCTCCACACCGCGGCACTTCGCCAGCCGCGC 6050
QY 1909 O-||| ||| ||| AlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyTh 1924
Db 6049 TCGACCTCGCGCGCAGGAGCGGAAACAGCGCCATCGCACCCACGACCCG-CCAGATAC 5993
QY 1924 rLeuAspGlyValTyrPro-||| ||| ||| ThrLeuMe 1933
Db 5992 CGGCCAATCAACCCCGCGCGAGCGAGCAGCAGCGCGCGCATCTCCAGCGACACGCGC 5933
QY 1933 tGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArg- 1952
Db 5932 CCGGCCACCGTGGCAGCAGCGATCTCCCTGGGAATGACCCACCCAGCCAGCCAGCTCC 5873
QY 1953 -AlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGl 1968
Db 5872 ACACCGATCGCTGCGCATGACCGCGCAGCAGCAGCATCACCGCCACCGCTACGAGGCT- 5815
QY 1968 yLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuValProVa 1988
Db 5814 -GCACCATCCACCGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 5765
QY 1988 lSer-||| ||| ||| GlyHisAlaThrIleAlaArgThrProAlaLysAsn- 2001
Db 5764 AAATCCCACTCCACATGAACACTCAGCGCGCGCGCAGCATCTCCCGCATCGACCGCGCAAC 5705
QY 2001 -||| ||| ||| 2001
Db 5704 ACCTCCGAAGACCCAGCAGCGCGCGCACCCATCCCGACCCCATCGCTGCCCTGACCGAGG 5645
QY 2001 -||| ||| ||| 2001
Db 5644 AAACCAACACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 5585
QY 2001 -||| ||| ||| 2001
Db 5584 ACCACACCGCGTCCGGATCACCCCGCGCAACCGCGCGCAACCGCGCGCGCGCGCGCG 5525
QY 2001 -||| ||| ||| 2001
Db 5524 TCACCCGAGCAGCAGCAACCGCGCGCGATACCAAAACCGCGAGCGCGCGCGCGCGCGCG 5465
QY 2002 -LeuAlaProHisAlaSerProProPro-||| ||| ||| 2012
Db 5464 GCCACATCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5405
QY 2013 -AlaProProAlaSerAlaSerAspPro-||| ||| ||| -Hi 2022
Db 5404 TCCCGCAGCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5345
QY 2022 sArg-||| ||| ||| GlnSerLysProPheSerIleGlnGluLeuGl 2037
Db 5344 TCAGAGCCGCGCAGCAGACTCCGACACTCAGAGCGGACCGGGGTCTGGGGCTCTCCGCGC 5285

QY 2037 uLeuArgSerLeuGlyTyrHisGlySerSer----- 2047
Db TCAGCGGTGTTCGGGGCTTCAGAGGATCAGATGGCGGTGGTACCACTGACACCAAAAGCG 5225
QY 2048 -----TyrSerProGluGlyValGluProValSerProValSerSerProSerLe 2064
Db GAGACCGCGCCGCGAGCGCGCGGTTCACCCCGCGGCATCCCGCGCTCGGTGACCAAC 5165
QY 2064 uThrHisAspLysGlyLeuProLysHisLeu-GluGluLeuAspLysSerHisLeuGluG 2084
Db CACACACACCGACCGACGAGTCCATCGCGCTCGGCTC-----ATCCATGCGAA- 5112
QY 2084 lyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaHisLeuP 2104
Db CCGTCCGCGCGCAACCA-----CCCATACCGCATCGCAACACCATCTTG 5069
QY 2104 roHisLeuArgProLeuProGluSerGlnProSerSerProLeuLeuGlnThrAlap 2124
Db ATCACAC-CAGCCACCCCGCAGCGCGCTGAGCATGACCAATATTCGACTTCAACGACCC 5010
QY 2124 roGlyValLysGlyHisGlnArgValThrLeuAlaGlnHisLeuValIle 2144
Db 5009 CA-----ACCA 5004
QY 2144 hrGlnAspTyrThrArgHisGlyProGlnGlnLeuSerAlaProLeuProAlaProLeu 2164
Db CACCGGAGCGCCACCATCAGATCCCGCCCATAGTCCCGCAGCAACCGCTGCGCTCGAT 4944
QY 2164 yrSerPheProGlyAlaSer-----CysProValLeuAspLeuArgArgProPro 2181
Db CGGATCACCCAGCGCGGTCCCGTCCCATCGGCTCCACCATCCACATCCGCGCCACCGT 4884
QY 2181 erAspLeuTyrLeuProPro-----AspHisGlyAlap 2193
Db CAACCGCGCATCCCAACCGCTTCGAGCAATCACCGCTCTGGAGGACCATTCGCGC 4824
QY 2193 roAlaArgGlySerProHisSerGluGly-----LysArgSerProGluProAsnL 2211
Db CGTCAACCATTCGACGACCGCTCTGATTCACCGCGCGCCACACACACCGCCCAAC 4764
QY 2211 ysThrSerValLeuGlyGlyGlyGluAspGlyleGluProValSerProProGlu----- 2229
Db CGGATGCCCAACCGCGCGGCGCATCCGACACCGCTCCACGACGACCGCCCGCACCTC 4704
QY 2230 -----GlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArg 2248
Db GCGCCCAAGCGGTCCATCCCGCCCATCCCGGAAC-GCC-----TTGACCGCC 4657
QY 2248 spGlyGluGlnThrGluProSerArgMetGlySerLysSer-----ProGlyAsnT 2265
Db CATCGCGCGCAACCCCGCTGCGCGAAACTCCAGACATCCCGCGCGCGCATCA 4597
QY 2265 hrSerGlnProProAlaPheSerLysLeuThrGluSerAsnSerAlaMetValLys 2285
Db CCGTCAACACCGCGG-----AACGCCAT-GTCACACTCG 4562
QY 2285 erLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArg----- 2299
Db CCACCGCGCAACCGCTGCGCGCGAGATGACGCCACCAACGACGACGACGCGGTA 4502
QY 2300 -----AsnGluProGluTyrAsnI 2306
Db TCACCGGTACGCGCGGACCTCCAGCCCGCGGTAGCGCCACACCGCCCGACACCA 4442
QY 2306 leSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMet 2326
Db CTCCCGACGTACCGGTGAGGAGGTAGCGCTGCTCCGCTCCGCTCGCTCGG-----G 4388
QY 2326 hrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIle 2346
Db AGCCGAGGTCCGTATTCCTCGCGCATCACCGCGCGGTAGACCGCGGTGCGGTGCG 4328

QY 2346 leArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaA 2366
Db AGCGAGT-GCG-----GCTCGATACCGGCC---GCTCGATCGCTCCCGAGG--- 4282
QY 2366 snAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaA 2386
Db -----TCTCCAGCAACAGCGCTGCTCGGATCATCGCCACCGCTCCACGCGCGA 4230
QY 2386 laAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLysValS 2406
Db -----ACTCCGCGG-----CATCGCGAAGAAC 4182
QY 2406 erGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAsp 2426
Db CCCTCCCGCACATACGATTCCTCCCGCACACCGCTCGGCTCGTACAAC--- 4129
QY 2426 rgProSerValSerValHisSerGluGlyAspCysAsnArgArgThrProLeuT 2446
Db -----CCTCCATCCAGCCACGATCACCGGGAACCGCGTCCACCCCT--- 4075
QY 2446 hrAsnArgValTyrGluAspArgProSerSerAlaGlySerThrPheProTyrAsnP 2466
Db -----CCACCAACCGCCACACGCT---CCGCGGAC 4041
QY 2466 roLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProPro-----ProProGlyL 2484
Db -----CCACCGCGGAACCGCACCCCATACCCACCGCATCGCTCTGGC 3987
QY 2484 euProAlaGlySerGlyProLeuAlaGly 2493
Db TTGGAAGCTCTTCTCGCGCGGTGCGAG 3958
RESULT 95
AE007164/c 15311 bp DNA linear BCT 27-APR-2001
LOCUS Mycobacterium tuberculosis CDC1551, section 250 of the
DEFINITION complete genome.
ACCESSION AE007164 AE000516
VERSION AE007164.1 GI:13883468
KEYWORDS Mycobacterium tuberculosis CDC1551
SOURCE Mycobacterium tuberculosis CDC1551
ORGANISM Mycobacterium tuberculosis CDC1551
REFERENCE 1 (bases 1 to 15311)
AUTHORS Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
TITLE Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 15311)
AUTHORS Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source 1. 15311
/organism="Mycobacterium tuberculosis CDC1551"
/mol_type="Genomic DNA"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="clinical strain"
complement(114..1662)
/gene="MT3613"
/note="This region contains an authentic frame shift and
gene

Db 8744 AGCAGACCGAGCGCGCGCATGATGATGCGGACACGTGACGAGCGGTTCT 8685
QY 1522 leAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeu 1542
Db 8684 GATTGCGCGGTGCG-GCTATGGCGCGCGGTGGGGATACCGTGGGTGCGCGTTG 8626
QY 1542 hrTyrGluAspHisGly-AlaProPheAlaGly- 1552
Db 8625 TCCTCAAACTCT--GGCTGACCACTGCGCGCGGACCGCGCGGCACCATCGCGCGC 8569
QY 1553 -----HisLeuProArgGlySerProValThr 1561
Db 8568 TGGGCCCATCGACATAGACCGCGCGCTCGCCCTTGCCTCGCGCGTCCGCGCG 8509
QY 1562 MetArgGluProThrProArgLeuGlnGlySerLeuSerSerSerLysAlaSerGln 1581
Db 8508 CCGCGCGCGGTGCCA-----CCGGTGGTGCCTTACCACCGTCTCCGGGCGCACCG 8455
QY 1582 AspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValPro 1601
Db 8454 CCGCTGCGGTGTCTCCCTTGGCG-----CCATTCCAGCGGTGCCG 8413
QY 1602 GluHisHisProHisProLysSerProTyrGluHisLeuArgGlyValSerGlyVal 1621
Db 8412 CCGTTCACCAACACCG----- 8395
QY 1622 AspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgLysIle 1641
Db 8394 -----CCGGTCCGCATCGCGCGGTGGCGGCCCGAGCGGTG 8356
QY 1642 ProLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAnProThrTyr 1661
Db 8355 CCACTGCG-----TTGCGCGCGGTGCCACCTTG 8326
QY 1662 ProHisLeuTyrPro-----ProTyrLeuIleArgGlyTyrProAspThrAla 1677
Db 8325 CCGCGGTGTGCGCGCGCGCGCACACCGCGTCCCGGGGCGACTCGGGCTCGGCG 8266
QY 1678 AlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHis 1697
Db 8266 ----- 8266
QY 1698 AsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGlu 1717
Db 8265 -----TCGCG----- 8260
QY 1718 SerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnVal 1737
Db 8259 -----CCATTGCGCGCGGTG 8245
QY 1738 ProHisLeuProValLeu-----ValProProThrProGlyThrProAlaThrAlaMet 1755
Db 8244 CCGCGCATACCGCGCGCGCGCGCGGTGCGCGTTCATGCTCCCGCGCGCGGTG 8185
QY 1756 AspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSer 1775
Db 8184 CCGCGGTGCGCATTCGCGCGCGCACCG----- 8155
QY 1776 ProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGlu 1795
Db 8154 CCGCTGCGCGCG-----GTGCGCGCGTCCGCGCGCGCATCGCGCGAC 8114
QY 1796 ArgGluArgAspArgAspArgGluArgAspArgGluArgGluLysSerIleLeu 1815
Db 8114 ----- 8114
QY 1816 ThrSerThrThrThrValGluHisAlaProIleThrArgProGlyThrGluGlnSerSer 1835
Db 8113 -----CCGGGTGCGCGCGCGCGCG 8093
QY 1836 GlySerSerGlySerGlyGlyGlySerSerSerArgProAlaSerHisSer 1855
Db 8092 GGTGCGCGGTGCGCGCGGTGCGCGCGCG-----GCCCGCGTCCCGCGCGCGGTGCC 8039

QY 1856 HisAlaHisGlnHisSerProLysSerProArgThrGlnAspAlaLeuGlnArgPro 1875
Db 8038 GCGGTGCGCGCATCGCGCGCATTCGCGCGCACCGCGGTGCGCGCGCGCGCGAC 7979
QY 1876 SerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaVal-GluProSerLysPr 1895
Db 7978 CTCGCGCGCTCGCGCGCGGTAC-----GCGTTGCGCGCGGTGCGCGCGCGCC 7925
QY 1895 o---ThrValLeuArgSerThrSerSerProValArgProAlaAlaThrPhePr 1914
Db 7924 AGTGTGCGCGCGACGCTCGCGCGCGCGCGCGCGGTGCGCGCGGTGCGCGGTGCC 7865
QY 1914 oProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetG1 1934
Db 7864 GCGGTGCGCGCGGTGCGCGCGCGGTGCGTCAGTAGTGGCGCGTGGCGCGCG--ATGGC 7808
QY 1934 uProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAs 1954
Db 7807 CCG-----TTGCGCGCATCGCGCGCGCGCGCTTGGCGCGCGGTGCGCGCGCC 7754
QY 1954 pThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSe 1974
Db 7753 AGCG-----GTCAAGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGAAGCC 7709
QY 1974 rProSerLysGly-----Se 1979
Db 7708 ACCTGCGCGCGGTGCGCGCGCGCGCGCGCGTGCAGTCATGTTTGGGTGTTATC 7649
QY 1979 rGluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAl 1999
Db 7648 TGCTCGGTGCGCGCGGTATCCGCGTCAGCGCGATGCCACCGTGCCTGCGCGGTGCCG 7589
QY 1999 aLysAsnLeuAlaProHisHisAlaSerProAspProProAlaProAlaSerAlaSe 2019
Db 7588 AGTACCGGTGAGCGCG-----CCGTGCGCGCGGTGCCCGCGTGCCTGCCGCGC 7541
QY 2019 rAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuAr 2039
Db 7540 GCGCGCG----- 7534
QY 2039 gSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerProVa 2059
Db 7533 -----TTCCGCGCG-----GCACCGCGGTGCCGCGT 7508
QY 2059 lSerSerProSerLeuThrHisAspLysGlyLeuPro-----LysHisLeuGluLe 2077
Db 7507 GCGGTGCGCGGTGTTGCCCGCATCGCGCGGTGCGCGCGGTGCCCGCGTGCCTGCCGCGTT 7448
QY 2077 uAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuG1 2097
Db 7447 GCGCGGTGCGCGTCCCGTCCCGGACCGCGGTGCGCGCGCATGCCGCG----- 7396
QY 2097 yGlyGluAlaAlaHisLeuProHisLeuArgProLeuPro----- 2110
Db 7395 -----ATGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCCGCGTGCCTGCCGCG 7343
QY 2111 -----GluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyVa 2126
Db 7342 GTTGGCGCGGTGCGCGCGGTGCGCGCGCGCGCGCGGTGCGCGCGGTGCCCGCGCGCGT 7283
QY 2126 lLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAs 2146
Db 7282 GCG----- 7279
QY 2146 pTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPh 2166
Db 7278 -----CCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7235
QY 2166 eProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuPr 2186
Db 7234 GCGCGCGCGCGGTGCCATTCGCGGATTAACGTTGCGCGCG-----TTACC 7187


```
QY 2186 oProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSe 2206
    |||||
Db 7186 GCATTG-----CCACTGTGCGCGGTGCGC----- 7159

QY 2206 rProGluProAsnLysThrSerValLeuGlyGlyGluAspGlyIleGluProValSe 2226
    |||||
Db 7158 -CGTGCCTG-----CGCTGCGCGCGTGCCTG----- 7133

QY 2226 rProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyProLeuLeuTy 2246
    |||||
Db 7132 TCACCC-----TGCCGCGCGCGCCACCCACCGCAGCGTTCACCGCT----- 7090

QY 2246 xArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSe 2266
    |||||
Db 7089 -----TGTCACCGAGCGCGCGGACCGTGCCACCGCCGACGCTGCG 7046

QY 2266 rGlnProProAlaPheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLy 2286
    |||||
Db 7045 GTCGCGCGCGGTG----- 7033

QY 2286 sLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyAsnIl 2306
    |||||
Db 7032 -----CCGCGCGCACCGCC----- 7019

QY 2306 eSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetTh 2326
    |||||
Db 7018 GTCGCGCGCGCCACCTGACCCCGGTCCGCCA----- 5982

QY 2326 rTyArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIl 2346
    |||||
Db 6982 ----- 5982

QY 2346 eArgLysAlaLeuMetGlyLysTyAspGlnTrpGluGluSerProProLeuSerAlaAs 2366
    |||||
Db 6981 -----CCGCTGCGCGCTTGCGCGCGCGCCACCTTCCTCCGCGCG 6941

QY 2366 nAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAl 2386
    |||||
Db 6940 GCGG-----CCGTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6888

QY 2386 a-----AspGlyArgSerAspHisThrLeuThrSerProGlyGlyG 2400
    |||||
Db 6887 CGCGGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6828

QY 2400 yGlyLysAla-----LysValSerGlyArgProProSerSerArgLysAlaLy 2415
    |||||
Db 6827 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6768

QY 2415 sSerProAlaProGly-----LeuAlaSerGlyAspArgProProSerVa 2430
    |||||
Db 6767 CTCGCGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6708

QY 2430 lSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTr 2450
    |||||
Db 6707 TGCGCGCGTGTGTAGTAGGAGGTCGCG-CGCGCGCT-CCGCGT----- 6664

QY 2450 pGluAspArgProSerSerAlaGlySerThrProPheProTyAsnProLeuIleMetAr 2470
    |||||
Db 6663 -----CCGCGCGCACCGCGCTTGCGCGCGCGTTCGCCAGCGCGTCCCGTTA----- 6622

QY 2470 gLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyPr 2490
    |||||
Db 6621 -----CGCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6587

QY 2490 oLeuAlaGlyPro 2494
    |||||
Db 6586 ACCGCGCACCGCG 6574
```

RESULT 96

AF357202

LOCUS

DEFINITION Streptomyces nodosus amphotericin biosynthetic gene cluster,
complete sequence.

113193 bp

DNA

linear

BCT 04-NOV-2002

```
ACCESSION AF357202
VERSION AF357202.1
KEYWORDS GI:14794889
SOURCE Streptomyces nodosus
ORGANISM Streptomyces nodosus
REFERENCE 1 (bases 1 to 113193)
AUTHORS Caffrey,P., Lynch,S., Flood,E., Finnian,S. and O'Liinnk,M.
TITLE Amphotericin biosynthesis in Streptomyces nodosus: deductions from
JOURNAL analysis of polyketide synthase and late genes
MEDLINE Chem. Biol. 8 (7), 713-723 (2001)
PUBMED 21344785
REFERENCE 11451671
AUTHORS 2 (bases 1 to 113193)
TITLE Caffrey,P., Lynch,S.V., Flood,E.M., Finnian,S.M. and O'Liinnk,M.
JOURNAL Direct Submission
SUBMITTED (07-MAR-2001) Industrial Microbiology, University College
Dublin, Belfield, Dublin, Ireland
FEATURES
    source
        1..113193
            /organism="Streptomyces nodosus"
            /mol_type="genomic DNA"
            /db_xref="taxon:40318"
            complement(4..1824)
            /gene="amphG"
            complement(4..1824)
            /gene="amphG"
            /note="possible amphotericin export protein"
            /codon_start=1
            /transl_table=11
            /product="AmphG"
            /protein_id="AAK73498.1"
            /db_xref="GI:14794890"
            /translation="MAPDVPPEHEEERESEQVRRRLAALLPHRRSVGLALPAGVVGI
            LNAFGKTLGRVTDLIAADGVLHGCGPAGVDGALGRLLMILLVLYVAVFVLMVQN
            WLAVSVLLIHLRHAQEKARLPFRDKPAGETLSRGTDVNDLQOQLTLT
            DLISVFSVLMISLMILISPLAGVLLSVPSGLLAATSKRAQPOVAAOWASGK
            LTAHVECMAGHALVKAEDRAEAEQFDERNEAVYRAGSGAQFAGALEPVMFVAN
            LGYAVAVGAWKVGSLTDGQAFILYARQSPQPIVETASVAGSQSGVASQRV
            FTLDAPQEPPEPRPLAVERVEGVEFDVDFRYPDPTLIEGLSLSEPGSTVAVV
            GSPGAKGTTVANLMRFYEDISGRLLDGTDTAANNRDLRSRGLVQDTWLFKGTI
            AENIAYGPGATRADIVAEARATYADRTLSQGYDTVLDDESGVAGEQLITVA
            RAFLARPAVLVDDEATSSVDTELLIQRAMNLTLPGRSTFVIAHRLSTIRDADVIV
            MKSGRIVSQGTHDQLIDAGQAYARLHAARADAPADVTVG"
            complement(1805..3628)
            /gene="amphH"
            complement(1805..3628)
            /gene="amphH"
            /note="possible amphotericin export protein"
            /codon_start=1
            /transl_table=11
            /product="AmphH"
            /protein_id="AAK73499.1"
            /db_xref="GI:14794891"
            /translation="MAPSVRWPTPFQRAAGRPQCHSVLLRLMRTQLRPVAGSVVA
            LVVHLVQLTLLPTLIGALLIDEGVVRHSDRIGTICATNAVVALVQIAALGAA
            LGARTALGRDLRSARFRRVLDPSAREIGRFTSLTRSVNDVQVQVNLQAGLGI
            FVAAPMCLGSLVLRALQDVTLLILVPMVLVAVCGLLLSRMAALVARLQOQLDRI
            GELLERITGVVRVSFARDAHEGRFTTNEELISGLSVGRLLIAMLPSVLLMNL
            FTLLGLWGARRIDSGSMQIGALSAFLSVLSILMSVVMIAFVFNLPRAVCABRIT
            EVLOAETDVPPASPRPWAGAGVAGVAGPRYPGAENAVLRDLSTLRPCERAVIL
            GSTGSKTLLHLILRVDTVAGVRIAGTVRELDPSVMAAAVQRPVLPAGTV
            ASNMRGPDATDEELWEVIRIAHQEHGFVRLRGGLDTEIAQGGTVSGGQRRLAIR
            ALLRRPATYLDFFDSFALDQSTEAARKALPYTEGATVITVAQRVASVRDADRIVLL
            DQGGIATGTHLDALLDSPTVREIALRSQRTRETAHGAGRS"
            3840..4874
            /gene="amphDIII"
            3840..4874
            /gene="amphDIII"
            /note="mycosamine biosynthesis protein"
            /codon_start=1
            /transl_table=11
```

```

/product="AmphdIII"
/protein_id="AAK73500.1"
/db_xref="GI:14794892"
/translation="MPKRALITGIGDGSYLAEHLISOGYQVWGLIRGOANPKFRV
SRLASLSFVDCDLMQGSLSVSDVQVYVNLGAI SFVPMKSWQAEHLVTEYNGMG
VLKLSAIRNWSGLSMRSTAGTSGQIRFQOASSEMFGRKAVTPQRETLFHFSPYIG
AAKAYGHF ITRNRYRESFGMYAVSMLFNHSEPRGQEFVTRKTLSDAVRAFKHGLQDKL
ALGNMADYVRYAGSYVRYAMHMLQDDAPDDYVIGTGBMTKSVARFAFVGLGDLW
KDYVYVDPDLVRPAEVEVLCASSKAQAQLGWKPSVDFQLMRMMVDADLASVSRQNE
LDLALLAHSW"
5042..33574
/genes="amphi"
5042..33574
/notes="amphi"
/genes="polyketide synthase multienzyme polypeptide housing
extension modules 9, 10, 11, 12, 13, and 14"
/codon_start=1
/translation_table=11
/product="Amphi"
/protein_id="AAK73501.1"
/db_xref="GI:14794893"
/translation="MNEOKLRDYLKATADLRARRRRVGELESASQEIATVGMTCR
YGGVSSPDLARMVEAGNGVTFPTDRGMDLEALASFTSRGGFLHDAPEPDADFF
GISPREAVAMPQQRVLESAMERAGINPTSVKSGRTGVFIGNAQOYRVRGPADG
AGFQUTNGTSGLSRI SYTFGTGPAVTNDTACSSSLVAHLTAQALRAEGLTAL
AGGVYMSGPGTF IEMRGGLSVDCGRCSFGDADGTGWAEGVGLVLERISDAIRN
GREILAVVRGTAVNODGASNGLTAPNGPSQOAVIEQALVNARSGDIDVFEAHGTGT
TLGDPVEAQLATYQOEDDEKPLLLGSVKSI SHTOAAAGVAGV KVMWNRHGV
PRTLADDETRHVDMSQGA VRILTENTEMPATCAPRAAVSFGISGTNAHTIVEQAP
EPEPDADDEPTAAVTGVLVPLVLSGRSPVIRQAQAALLTLTGTFPPAPADUAY
SLATTATFEHRAVLASLDLPETLGLRTAI AEGTDPAVLADTVTGTDRETLRLVFT
GOQAQLGARELAEFPVFMAALDAALDAFTPLHDPVLRKVLWGSDADRLDRTETAYQ
PALFAVEVALYRLLESFVEKPDHLAGHSGEIAAAHVAGVFLSDDAATLVAARGLMQ
ALPEGMAVAVQSEDEVAFLLAGHEDLVSLAANGPSAVLSGDETTTVEALRLAA
DGRKTRSLRVSHAFBPLKAPMLDFRNVEGLTLSPILLPVVSDVTGEPATVAQLTS
PDYVDPHVRQFAATQKDWLKHVDVTFLELGPDSVLSMAQNCLODASDRLATVPA
LREGRPDTFTTAAALHTQGTALHDACFTGTGARTDLSLPTTYTQRRRYPRAVQG
GAADLSVGLAGHHPHLLSAAVSLADSEGLITGRI SLLSHPNLADTVRGATLLPGT
AFLELAVRAGDEVCGRVDELTAAPLVLPQEOGVQVQLWIGNPDASGRSSTVYGRP
DADEADPMTSHANGVLSASRTTSDPDATWPPADAEPLVDGLYERLASRGGVGLP
QGLRAWRGDEVAEVLVLPESGTDASFGFLHALLDSALHNASFVLDLERAAGSLP
FSWEGVSHASGATTLURLAPAGDAVAIAVADDSGLVLSADSLILRAVAREIDA
AAALVRDALFRLDMPVPTVAASGTAAALVGEDPGLRALPQDGLVAHPDLADLAA
DGAVPDTLLPTLGTGDPADPVTAAHRAATEALAAVRLWEQDERFAASRLALVTGA
TTCHDPAARAAVFLVSAQSENPGRFLVLDLADODTPALPAALTSBEPOLAVRGE
LEARLVRPASTAEAVAPRAGGEGAVLTGTTGGLGAVLASHLVAEHGVLLVSR
GMAAGAEVLAELESARATVACDVTDRAAVAELVAHPVSAVHSAAGVLDGNNV
GTLTPRLTTLVRPKVDAAWNLHEATRDLDLKAFVLFSSVAGVLSGPGQANVAGNAF
LDALAAHRAAGLPLGLSLANGPWEQGTGTTGTTSEDDLRRMPAGTTPALTVEQGLAL
DIALDQDDEALAPVRLDLSVLRAGQVPPILRLIRGRSRAAVAGSAGTAGGLAORLA
RLDAESGRDELVLDRGOVALVGHATGAEIDAGRAPRELGPDSLTAVELRNLNTVT
GLRLPATLPDPTVTSLSASYLDLGTTEVAEVVQRTAAVADDPITVIMWACHYCP
GGVTSDDLURLVTEGTDVSGPPNRRGMDVENLHYDPDHPGCTATRSGGFLHEAGE
FDPFGFGMSRLATDSQORLLLEASWEAIEHAGIDPVLGRSAGTGVFAGVWYSDYS
AMLSGPEFGVSGSGSSPLASGRSYTLLEGPAVTVDPTACSSSLVAMHMAQMLRS
GEISLAGAGGVWMSPTPVPFVDFARQGLSPGCRCKAFSDAGSAGVSGVGMVQLER
QSDAIRNGHOLLAVRGSVAVNODGASNGLTAPNGPSQOAVIRQALASGLSGLDVDDV
EAHGTCTTGLDPLLEAQLLATYGRDRDPEQPLLLGSVKSI GHTQAAAGVAGVIMVMV
SHRGVLPRTLHVDASSHVDWTEGAVELLTEQTAMPETGRPRRAVVSFGISGTNVH
TVLEQAPTTVPAPAPERTAPVPLLLSGTRDRAQAARALLTHQHPSPSLADL
GHISLATTRFRERFAVIAIQDREGLLASLSLAAGRPDPAVVEGEAAGHAAVAFMSG
QSQRAAMGRDELVTQPRFAAFDEVCALDPLLDRLPLREVPFAESGSEALLLDRTG
WTQPALFAVEVALYRLVESGWVRADFTVTHSICEIAAAHTAGVFTLQDARLVAARAT
LMEALPSGMVAVQATEBEVAPLLGEGLSVAVANGPTSVVSGDDEDAVELAAEFSG
RGRTKRLRVSHAFBSPHMDMLDAPRTVAETLSYAAPRIPLVSDLTGRADDAEVRT
ADYVVRHREAVFADCVTRLDRDAGTFLLELGPDSYAAPRIPLVSDLTGRADDAEVRT
PLLRADRPBESAAATAAQHIOHVDLMDWTAYLAGTCARRVOLPTTFAVSHRWNQPOK
SAPSPAGDPADQKLAWAVERGDAAELAAVLGDESLTLPDLSLPAALSSWRNNGOEK
ALLDLYRVWETRLSKPAAPVLDGNTWLLVSSDATADDETLDDGLAEALGAHGAVRV
RLVLDCADRDLARLADTENADNTAQLVLSVLPDREPTDGPAGFTOGLALTIALV
QALUADTGAGHRLMTATRGAVSTGPDTPHQAQTAWGMGRGVALSHPRLWGLVDLP
AFDRGAGQRLAEVLAKQADGEDQVALRATGVGRRLVRHIVDLSLPQARTASG
VLTITGTTGGIGAETARLARSAAHLVLTSSRRGPDAPGAELAELEQSGASVSIAC

```

gene

CDS

Alignment Scores:

Pred. No.:	0.000188	Length:	113193
Score:	654.50	Matches:	644
Percent Similarity:	30.30%	Conservative:	283
Best Local Similarity:	21.05%	Mismatches:	1086
Query Match:	4.95%	Indels:	1060
DB:	1	Gaps:	140

US-09-522-753-5 (1-2517) x AF357202 (1-113193)

QY	39	LeuLeuGluTyrGlnHisSerArgAspTyr-----AlaSerHis---52
DB	11581	CTCTTCGCATCAGCGGCACCAATGTCACACCGCTCTGGACAGCGCGCGGCACAC 11640
QY	53	-----LeuSerProGlySerIleGln60
DB	11641	CGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11700
QY	61	ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAanGluArgSerGln80
DB	11701	CGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT---11742

DVADRALAAVLDSLADQPLTGVVHTAGVGHYGLDALTPAEFAGLTAAKLAGAAHL
 DNLLGDRLEDFILFGSIAGVMSGDSQAYANAYDALARRARRAGLAATIAWGP
 WLLGTGAADAVSGTIRROGLGLDPALTEMRRAVVRQDVTVTVADVDDVYAPLF
 TARSALISDLPEVRALAEENTPATDGTASEIVORVRSISEPEQLRLITDLVTRAA
 TVLGHSSAGAVEDGAFRIGPDSLTAVELRHGLGATGLSLPSTWFDVPTPLELAQ
 TYRAEWSVLEVAGPAGTGGDDDEFIAGMSCRYPGSSPEQWDLVLSGTDAIT
 DPFVNRGNTAGLYDDPDHPGTSTYTGGLFHEADFDFMFGLISPRALVNDPQOR
 LLETWEAFERAGLTPTDLRGLSTGTFIGSSYQYQYGMGADGAEGLVTGTSPLVS
 GRLAYVFLGEPAVTDTACSSSLVALHLACQALRNGESNLAVAGCATVMTTSPVFA
 FQRALADGRCFKAESADGMTLAEGVGI VLVERLSDRNRNGHPLVAVIRGSAINO
 DGSNGLSAPNGSPQORVIRQALANARVAPGEIDLEAHGTGTPLDGTAEQAALFATY
 GRITPETALLGSGVKSNGHSQASAGVASIKWVWALRHGVMPQITLHADEPSSHDW
 SPGTVRLGENTDMPTGRPRAAVSSFGISGTNAHVILEQETEAFAADEQLAPAPL
 PVAAGVWPWLLSARGAALREQADRLTLHTVADPAARTIDIGLSLATRALFEHRAV
 VYPAGTODLEALRAVADGPGVARGVADVAGRTVFPFGQSGWAGQALLDVES
 PVFERRIASCALAEFTDWNLDLVRGAEGAPTLERVVDVOPASVAMVSLAAWRA
 QGVEDAVVGHSGQEIATAAVSGLSLRSGARVVTLRAQAI GRSLAGRGMMSVLAPV
 AEVEARKEAFEGRVSAANPRSSVAVAGEPEALDELHAQLTAETERRARVADYASH
 SHVEDLDEHLEILAEAVPTSEIPTFTVTGMDLDTTVMQDGYWYRSLRGLVLFAD
 AVRDLTADHRAPIEVSSHPVLAMSQMDIDDAGVAGVSGTLRRDNGGLDRFLLSAA
 EVFVRGVQVQWAAVFEFGTASRVOLPTYAFQHENLWMAAAPEAVTAADPEDAFWTA
 VEDGVSAALTAALGTDEDSAAVLPALSSRRARKERTVDSWRYPTKPTKLPQR
 TLDGTWLLVSADGVDDTVAELETGGAERVLVLDESCTDRAVLPERLTDADGLGI
 VSVLAGERTIGAVPGTGLVGLVVALTVVALQALGDAGIDTPLWALTRGAVTSRSDKVT
 APVQOVTGIGMTALECPGRWGVVDLPETLDDARAGQALAAVLALGALGDDDDQIALRS
 SGVFTERRIVRAADAPDGSARDWKPRGTTLTGSGTTLAPHLARLAEQGAELHVLVS
 RQPEAPGAELRAELAEARTETTLAACDITDRDAVLALESKAEGRTVTVVHTAAT
 IELHTLDTLDDFDRVLAAKVTGAOILDELDDDELDVLYSSTTGWGMSGAHAAY
 VAGNVALAELAEHRRARGLTALSLSNGIWAADLOLGRVDPMQIRRSGLREMPDQALS
 GUKAULDDQEVIAVDWETVHTVPTPTLDFEVEVQRLTAARAEQSGADPAR
 GFFAALLAPAAEQHRLKLETVTEAASVGLSSAEDLTDQAFRDRVGFDSLUTAVGL
 RNRLASVLTLPSTVFPNPAALAGFLHSELADHAGAVTAGVADPDDPTAI
 VMSCRYPGCIITSAEOLMRELVESVADSVFPADRGWDAEALYDPDASGRSTYVQG
 GLRDVADPDFGFGI SPREALSMDPQORLLLETAWEFENAGLDPVAQGRSTYTFI
 GASQDYGAAPVSGSESGHMITGSLPSVLSGRVSYLFGLEGPVTLDTACSSSLVAI
 HLAQCSLNGESTLALAGSINSPMSFGRQALAEADGCKKAYAGADGWTLAE
 GYGLILLERLSDARRNGHEVLAVIRGSVNGQASNGLTAPNGPSQOAVIRQALANAG
 VASANDIVLEHGHTGTALGDTIEAQALFATYQKDRDPERPVLISYKSNIGHTQMASG
 FEANIKLHALREGVAPKSLHIDQPSHTVDSGSGTQLLTERTEWETGRPRRAAVSS
 FGLSGTNVHTVLEQAPADAPAEADTPADALVPVLVSGRGAALRAQAQALGLDILA
 ERPGHTPLDLAFSLATSRALAEHRAAVADHDHEALVRLGTALRDGLPGALVQGRTR
 GRTATLQGGSRGRLRELYRVPFVADALDVLARIDGTTTERPDLVFAABESQ
 DAALLHRTGQAPALFALFLEVALFRLESWGVTDPDLAGHSVGEIAAAHVSRVVDLDDA
 CTLVARGMLQALPEGGMVALAEAEDEVLPLEGLAQDQSVAAHVAGSVVYVADGE
 EPLVALAAPEQGRRTKRLRVSHAFBPLMDPMLDPAFAVRLTHAPSLIPFVSNV
 TCTLAAPEOVCTADYVWVSHRVAEPADGIGLWSTOGGVQTFLELPGDGLVSPWARES
 LTDASTALLPTLRDRPEEQALVTAVAHAHGFVDVNTAMPQSGARRVALPTYAF
 QWERYWPTTAAGITAPAGSALDAEFMAAHEADVASLUTASLGLDDATVYAWVPEA
 AWORRGEQSLDSWRYRVTRKPGGAGAAPTGRWLVLVPAEHRDEATAAWAADVEA
 ALATATVRVEVTGTDRAALARLTAADGDTFGQLSLILALAPGDAGHPGAPALTLT
 ATALQALGADARIDAPLWNTIRGAVAVGRSEQVTAPBQAAVWGLFRAALAELEPARVGS

QY 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
Db 11743 -----CACCCACTCCAGAACACACCCCGAGCGCTCCCT----- 11775
QY 101 GluPheileGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
Db 11776 -----CGCCGACCTCGG-----CCACTCCCTGGCCACCCCG 11808
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysArgSer 140
Db 11809 CTCCTCGCTTCAGCGCGCGCGCG-----GGTGATCGCGCA 11844
QY 141 LeuThrGlyLys---LeuGluProValSerProProSerProPro-----HisThrAsp 157
Db 11845 GG-ACCGCGAAGACTGCTCGCTTCCTCGGCTCCCTCGCGCGCGCGCGCGCGCG 11903
QY 158 ProGluLeuGluLeuValProProArgLeu-----SerLys 169
Db 11904 CCGTCGTCGAGGTGAGCGCGCGCGCGCGCGCGGTGCGGCGGTGCTTCCTCGCGCGAG 11963
QY 170 -GluGluLeuileGlnAsnMetAspArgValAspArgGluileThrMetValGluGln 189
Db 11964 GCAGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12023
QY 189 nileSerLys-----Le 193
Db 12024 CGTTCAGCAGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12083
QY 193 uLysLysLysGlnGlnLeuGluGluAlaLysPro----- 207
Db 12084 TGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12143
QY 208 -----ProGluProGluLysProValSerProProL 219
Db 12144 CCGCGCTGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12203
QY 219 eGluSerLysHisArgSerLeuValGlnleileTyrAspGluAsnArgLysLysAlaG 239
Db 12204 CCAGCTTCGTACCGGTCACTC-----CATCGGTGAGTCGCGCG 12242
QY 239 uAlaAlaHisArgileLeuGluGlyProGlnValGluLeuProLeuTyrAsnG 259
Db 12243 CCGCGCACATCGC-----CGGGGTGTTCACCTCCAGGAGCGCGCGCGCGCGCG 12296
QY 259 nProSerAspThrArgGlnTyrHisGluAsnleileLysleAsnGlnAlaMetArgLys 279
Db 12297 CCG-----CGCACCTCATGAGCGCGCTGCGCGT----- 12327
QY 279 sLeuileLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCy 299
Db 12328 -----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12360
QY 299 sGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgileGluAsnAs 319
Db 12361 -----GGAAG 12365
QY 319 nProArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProG 339
Db 12366 AGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12402
QY 339 uileArgLysGlnArg---GluLeuGlnGluArgMetGlnSerArgValGlyGlnArgG 358
Db 12403 ----CGCGGTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12458
QY 358 ySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluileleAspG 378
Db 12459 AGCTCCCGCGCGGTTCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12516
QY 378 yLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValileProProMe 398
Db 12517 TGCCTTCCACTCGCGCGCATGAGCCCGATGCTTCGAGCGGTTCGCGCGCGCGCGCG 12576

QY 398 tLeuTyrAspAlaAspGlnGlnArgileLysPheile---AsnMetAsnGlyLeuMetAl 417
Db 12577 GCTGTCTTACGCC---GCCCGCGCATCCGCTGTCTCCGACCTGACGGCGCGAGGGC 12633
QY 417 aAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGlu 437
Db 12634 CGACGACCGCGAGGTG-----CGCACCGCGACTACTGGGTGCGCCACGTCGCG 12681
QY 437 sGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuileAlaSerPh 457
Db 12682 TGAGGCGGTACGG----- 12694
QY 457 eLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnG 477
Db 12695 -----TTCGCGCGACTGCGTCG-CACCCCTGCGCGCGCGCGCGCGCGCGCG 12737
QY 477 uAsnTyrLysSerLeuValArgSerTyrArgArgArgGlyLysSerGlnGlnGln 497
Db 12738 TCTTCTCGAAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12797
QY 497 nGlnGlnGlnGlnGln----- 503
Db 12798 ACGAGCGTTTACGACACACACACGCGACTCGTCCGCTGTGCGCGCGCGCGCGCGCG 12857
QY 504 ----GlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluLysAspG 522
Db 12858 AGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12917
QY 522 uLysGluLysGluLysGluAlaGluLysGluGluLysProGluValGluAsnAsp 542
Db 12918 CCGCGTACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12977
QY 542 sGluAspLeuLeuLysGluLysThrAspThrSerGlyGluAspAsnAspGluLysG 562
Db 12978 ACGCCCACTACTG----- 12990
QY 562 uAlaValAlaSer-----LysGlyArgLysThrAlaAsnSerGln----- 575
Db 12991 -CGCGCAGTGCCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13049
QY 576 -GlyArgArgLysGlyArgileThrArg-SerMetAlaAsnGluAlaAsn---SerGluG 594
Db 13050 GGGCGCGCGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13109
QY 594 luAlaileThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuLeuAsnGluSerS 614
Db 13110 ACAGCTCACCCCGTGAAGTCCCTG-----CTGCGCGCGCGTGAAGCT 13151
QY 614 rGTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsn 634
Db 13152 CTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13211
QY 634 rPserAlaileAlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheTyrP 654
Db 13212 GAGCGCGGTGAGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13242
QY 654 heAsnTyrLysLysArgGlnAsnLeuAspGluileLeuGlnGlnHisLysLysMetG 674
Db 13243 -----CCTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13294
QY 674 luLysGlu-----ArgAsnAlaArgLysLysLysLysLysAlaProAlaA 689
Db 13295 ACCGAGCTGCTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13353
QY 689 laAlaSerGluGluAlaAlaPheProProValValGluAspGluMetGluAlaSerG 709
Db 13354 CCGTCTCC-----TGGACGCGCGACT----- 13373
QY 709 lyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyA 729
Db 13373 ----- 13373
QY 729 sGlnValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThr 749

Qy	1265	yeGlyHisVal-----IleTyrGluGlyLysLysGlyHisValL 1278
Db	15399	AGSGCCACCTGGTACCGGCACCGCCACGAGTGCTCTCCGGCGCGCTCGCCTATGTCT 15458
Qy	1278	euSerTyrGluGly----GlyMetSerVal----ThrGlnCysSerLysGluAspGlyArgS 1296
Db	15459	TCGGCTGGAGGCGCGGGGTACCGTGCACACCGCTGCTCGCTCTCGCTCGCTCGCTCG 15517
Qy	1296	erSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluG 1316
Db	15518	CTGCACCTGGCCCTGGCCA-----GGCGCTGCCCAACGGCGAGAGCAACTCGCGGTGCGG 15571
Qy	1316	lyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleP 1336
Db	15572	GGCGGCGCCACCGTCATGACCACCCCAACGGGTTCGTGGCGTTCAGCGCGCAGCGCGCC 15631
Qy	1336	ro-----ProGluArgHisSerProHisH 1344
Db	15632	CTCGCAGAGACGGCGCGGTCAAGCGGTTCTCCGAGAGCGCGCAGCGCATACCTCGCC 15691
Qy	1344	isLeuLysGluGlnHisHisIleArgGlySerIleThrClnGlyIleProArgSerTyrV 1364
Db	15692	GA---AGGTGTCCGCATGCTCTCTGTCGAGCGGCTGTCGACGCCCGCGCAACGGCCAC 15748
Qy	1364	alGluAlaGlnGluAsp---TyrLeuArgArgGlu---AlaLysLeuLeuLysArgGluG 1382
Db	15749	CCGGTGCTCGCGGTGATCCGGGCTCCGCATCAACGAGGAGGGCGCTCCACGGCCCTC 15808
Qy	1382	lyThrPro-----ProProProPro----- 1388
Db	15809	AGCGCCCCAAACGGGCCCTCCACGAGCGCGTCATCCGCGCAGCGCCTGGCCAAACGCCCGG 15868
Qy	1389	-----ProSerArgAspLeuThrGluAlaTyrLysThrClnAlaLeuGlyProLeuLysL 1407
Db	15869	GTCGCCCGCGCGAGATGCACCTCTCGA-----GCCCCACGGGCACC 15910
Qy	1407	eulysProAla-----HisGluGlyLeuValAlaThrValLysGluAlaGly---- 1422
Db	15911	GGCACCCCGCTCGGCACCCCATCGAGC-----GCAGGCCCTG 15949
Qy	1423	--ArgSerIle-----HisGluIleProArgGlu-----GluL 1433
Db	15950	TTCGCCACCTACGGCGCGCACCGGTACCCCGAGACCGGCTGCTCTCGGCTCGGTGAAG 16009
Qy	1433	euArgHisThrProGluLeuProLeu----- 1441
Db	16010	TCGAACATCGG---GCATCTCCAGTCCGCGCGGTGTCGGAGATCATCAAGATGTCTC 16066
Qy	1442	-----AlaProArgProLeuLysGluGlySerIleThrGlnG 1454
Db	16067	ATGGCGCTGGCCACCGGTGTGATGCCCCAGACCTTCGACGCCGACGAGCCCTC----- 16119
Qy	1454	lyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspVal- 1473
Db	16120	--CAGCCACGTGCAGCTGGAGCCGGCACCGTACCGTCTCGGCGGAGAACACCGACTGG 16177
Qy	1474	-----ArgSerLeuIleGlySerProGlyArgThrPheP 1485
Db	16178	CCGCAGACGGCGCTCCGGCGCGCGGTGTCTCTCTCGGCATCAGCGGTACCAAC 16237
Qy	1485	roProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrG 1505
Db	16238	GCCACGTATCTCT-----GGACGAGGACGACGAGCGGCCCGCC 16276
Qy	1505	luGluSerLeuLysSerArgProGlyThrAlaSerSerSerGly-----GlySerIleA 1523
Db	16277	GCCGAGGACGAGCAGCTCGCCCGCGCGCCCTGCCGTCGCGCGCGGTGTGTGTCGCGTGG 16336
Qy	1523	laArgGlyAlaPro-----ValIleValProGluLeuGlyLysProArgGlnSerProL 1541
Db	16337	CTGCTGTCCGCCGTGGCGCGCGCTGCGCGAGCAGCGCGCGCTCTCTCACCCAT 16396

Qy	1541	euThrTyrGluAspHisGlyAlaPro-----PheAlaGlyHisLeuP	1555
Db	16397	CTCGTGACCGCCGACCGCGCGCCGCGATCGACATCGGTCTCTCCCTGGCCACCTCC	16456
Qy	1555	roArgGlySer-----ProValThrMetA	1563
Db	16457	CGCGCGC-TCCTCGAGCACCGCGCGTGTCTGTCGCCCGCGCACCGCCGCTGGA	16515
Qy	1563	rgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerLySAlaSerGlnAspA	1583
Db	16516	GGCGCTCGCGCGCTCGCGCAGACGGGCGCTCGGCGCTGCTCCCGCGCA	16565
Qy	1583	rgLysLeuThrSerThrProArgGluLeuAlaLysSerProHisSerThrValProGluH	1603
Db	16566	-----GTGTCGCGGACGTCGCGGCGCGCACCGTCTTCGTTCTCCCGCGCA	16611
Qy	1603	iHisProHisProLleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspL	1623
Db	16612	GGGCTCCC-----AGTGG-----	16625
Qy	1623	euTyrArgSerHisLleProLeuAlaPhe--AspProThrSerileProArgGlyLleP	1642
Db	16626	-----CGGGATGGGCGCCAGCTCTCTGACGAGTCCCGGTGTTCGCGAGCGATCGC	16680
Qy	1642	roLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThr-Tyr	1661
Db	16681	CG-----AATGCCCGCCGCCCTCGCGGAGTTTCAACGATCGAA	16719
Qy	1662	ProHisLeuTyrProProTyrLeuLleArgGlyTyrProAspThrAlaAlaLeuGluAsn	1681
Db	16720	CCTCATCGAGCTCTCGGGCGC-----CGAGGGCGCCCGACGCTGGAGCGCGTGCAGCT	16776
Qy	1682	Arg-----GlnThrileAsnAspTyrLleThrSerGlnGlnMetHisAsnThr	1699
Db	16777	CGTCCAGCCGCTCGTCTCGCGTGAATGTCTCTCTGGCGCGCTGTG---CGGGGCCCA	16833
Qy	1700	AlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSer	1719
Db	16834	GGGTGTCGAACCGACGCGGTGTCTCGGCACCTCGAGGCGCAGATCGCGCGCGTGTGT	16893
Qy	1720	LeuAlaLeuAsnTyrAlaAlaGlyProArgGlyLleLleAspLeuSerGlnValProHis	1739
Db	16894	CTCGCGCGCTGTCTCGTCTCGCGACGGCGCGCGGTGTGACGCTCCGCGCCGACGCAT	16953
Qy	1740	LeuProValLeuValPro-----ProThrPro---Thr	1748
Db	16954	CGGCGCA-CCCTCGCGCGCGCGCGGCGGCGGATGATGCGGTGGCGTCCGCTGGCGAGG	17012
Qy	1749	-----GlyThrProAlaThrAlaMetAspArgLeu	1758
Db	17013	TGGAGCGCGCTCGAGGCGTTTCGAGGACGGGTGTCTCGTCTCGCCCGCAGA-----	17063
Qy	1759	AlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSer-----	1773
Db	17064	-----ACGCGCCCGCTCTCTCGTCTGTCTCGCGCGCAGCCCGAGCGCCTGGACG	17111
Qy	1774	-----SerSerProLeuSerProGlyGlyPro-----Thr	1783
Db	17112	AATGTCGACGACAGCTCACCGCGAGGATCGGGCCCGCGGTGCGCTGGACTAG	17171
Qy	1784	HisLeuThrLysPro-----ThrThrThrSerSerSerGluArgGluArg	1798
Db	17172	CCTCGCACTCCCCCATGTGAGGACCTCGACGACGAGATCTCGAACTGCTGGCCGAGG	17231
Qy	1799	AspArgAspArgGluArgAspArgGluArgGluLysSerileLeuThrSerThr	1818
Db	17232	TCGCGC-----CGCGGACGCTCGAGATCCCGTCTCTCTCGACCGCTCACCG	17276
Qy	1819	ThrThrValGluHisAlaProIleTyrArgProGlyThrGluGlnSerSerGlySerSer	1838
Db	17277	GGCACTGGCTCGACACACCGTCATGAGACCGCGGTACT-----	17315
Qy	1839	GlySerSerGlyGlyGlyGlyGlySerSerSerArg-----	1850

Db 17316 GGTACCGCAGCCTGCGCGCGGTGCTGTTGCGGACGAGTCCGTCGACCTGATCCCG 17375
Qy 1851 -----ProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArg 1866
Db 17376 CCGACCAACCGTGCCTTCATCGAGGTACGTCACCCCGGTGCTCGCATGTCGTCGTCAGG 17435
Qy 1867 ThrGlnAspAlaLeuGln-----ArgProSerValLeuHisAsnThrGly 1882
Db 17436 ACATGATCGACGACCGCGGAGTGGCGGTGTCGCTCCGCGACCTCGCGCGGACACG 17495
Qy 1883 MetLysGlyIleThrAlaValGluProSerLys-----ProThrValLeuArgSer 1900
Db 17496 GCGGTCTCGACCGCTTCTCTGCTGTCGCCCGCGAGGTGTCGTCGCGGTGTCAGGTG 17555
Qy 1901 ThrSer-----ThrSerSerProValArgProAlaAlaThrPheProAlaAlaThr 1917
Db 17556 ACTGGGCGCGGTGTCGAGGGACCGTGGTGGTCCCGGTGCGACCTCCACCT----- 17609
Qy 1918 HisCysProLeuGlyThrLeuAspGlyValThrProThrLeuMetGluProValLeu 1937
Db 17610 ---ATGCTTCCAGCAGCAACCTGTGGCCATGGCGCGCGCGCGCGGCGGCGGCGG 17660
Qy 1938 LeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHis 1957
Db 17661 TCACCGCGCGCGACCGCGAGGACCGCGCT----- 17690
Qy 1958 AlaPheLeuAlaLysProProAlaAlaArgSerGlyLeuGluPro-----AlaSer 1973
Db 17691 -----TCTGACCGCGCGTTCGAGCGGAGGACGTCTCCGCGCTGACCGCGCGCGCTCG 17741
Qy 1974 SerProSerLysGlySerGluProArg-----ProLeuValProValSerGlyHis 1991
Db 17742 GCACGACGAGGACTCGTCCGCGCGGTGTCGCGCGGTGCTCTCTGCGCGCGGCGCC 17801
Qy 1992 AlaThrIleAla-----ArgThrProAlaLysAsnLeuAlaProHisHisAlaSerPro 2009
Db 17802 GCAAGGAGCGCTCCACCGTGCAGTCTCTGGCGCTACCGCGCGCTGGAAGCCGCTCACCA 17861
Qy 2010 AspProProAlaProProAlaSerAla-----SerAspProHisArgGluLys 2025
Db 17862 AGTCCCGGACGACCGCTCGAGCGGACCTGGCGGACCTGGTGTGTCGCGCGCGGTGCGAG 17921
Qy 2026 ThrGlnSerLysProPheSerIleGlnGluLeuArgSerLeuGlyThrHisGly 2045
Db 17922 ACACGAGCGTCCGAGGCGCTGAGACGCGCGCGCGCGCGGCGGCGGCGGCGGCGG 17972
Qy 2046 SerSerTyr----- 2048
Db 17973 TCCTCTGGACGAGGCTGCACCGACCGCGCGGTGTCGCGGAGCGCTCACCGACGCGCG 18032
Qy 2049 -----SerProGluGlyValGluProValSerProValSerPro 2062
Db 18033 ACGGCTGACCGGATCGTCTCGTCTCGCGGCGCGCGGACCGGTGCGGTCCCG 18092
Qy 2063 SerLeuThrHisAspLysGlyLeuProLysHisLeuGluLeuAspLysSer----- 2080
Db 18093 GCACCGGCTCGTCTCGGCGTCCCTCACCGTCCGCGCTCGTCCAGGCGCTCGGCGAG 18152
Qy 2080 ----- 2080
Db 18153 CCGGTATCGACACCGCGCTGTGGCGCTGACCGCGCGCGCGCTCTCCACCGCGCGTCCG 18212
Qy 2081 ---HisLeuGluGlyGluLeuArgProLysGluPro-----GlyProValLys--- 2095
Db 18213 ACAAGGTACCGCGCGCTCCAGCGCGGACCGGTACCGGATCGGTGACCGCGCGGTG 18272
Qy 2096 -----LeuGlyGlyGluAlaAlaHisLeuPro---HisLeuArgProLeu--- 2109
Db 18273 AGTCCCGGACGCTGGGCGGTGTCGTCACCTGCGCGGACGACACTGGAGCGCGCGCGG 18332
Qy 2110 -----ProGlu 2111
|||

Db 18333 GCCAGCGGCTCGCGCGCGTCTCTCGCGCGCGGTGTCGCGGACGAGCACCAGATCCCTCGC 18392
Qy 2112 SerGlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArg 2131
Db 18393 GCTCTCCGCGGCTTTCACCGCGGTATCTCGCGCGCG----- 18431
Qy 2132 ValValThr-LeuAlaGlnHisLeuGluValIleThrGlnAspTyrThrArgHisHis 2151
Db 18432 -----ACGCGCGCGCGCGGTCTCGCGCGCGACTG-----GAAGCGCGCGCGACCA 18479
Qy 2151 sProGlnGlnLeuSerAlaProLeuProAlaPro----- 2162
Db 18480 CCCTGTGTACCG-GCGGCTCCGCGACCTCGCCCGCGACCTGCGCGCGCTGCTCGCGCGAG 18538
Qy 2163 -----LeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgAr 2178
Db 18539 CAGGCGCGCGAACACCTGTGTCTCGGTGACGCGCGCGCGCGCGCG-----AGCGC 18586
Qy 2178 gProProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerPr 2198
Db 18587 CCGCGCGCGCGCAACTCCGCGCGCGAGTGGCGGACCGCGGACCGAGACGACCTCGCC 18646
Qy 2198 oHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerVal-LeuGlyGly 2218
Db 18647 GCTGTGACATCACCGACCGGTGACCGCGCTCGCGCGCTCTCGAAAGCCTGAAGCGCGAG 18706
Qy 2218 LysGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerA 2238
Db 18707 GCGCG-----CACGCTCCGCGCGGTGTGTCACCGCGCGCGCACCATC 18748
Qy 2238 rgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetG 2258
Db 18749 GAACCTGCA-----CACCTCGACGCGCA-C 18771
Qy 2258 LysSerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluS 2278
Db 18772 CACCTCGACGACTTCGACCGCGCTCTCGCGCGCAAGGTCCACCGCGCGGACGATCTCGA 18831
Qy 2278 erAsnSerAlaMetValLysSerLysLysGlnGluLeuAsnLysLeuAsnThrHisA 2298
Db 18832 CGAATCTCTCGACGAGGAGC-----TGACGACTTCGT 18867
Qy 2298 snArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProA 2318
Db 18868 CCGTACTCTCTCA-----CCGC 18885
Qy 2318 laIleThrGlyThrGlyLeuMetThrTyrArg---SerGlnAlaValGlnGluHisAlas 2337
Db 18886 CGGCATGTGGGCGAGCGCGCGCGCGCGGTATGTGCGCGGCAACG----- 18932
Qy 2337 erThrAsnMetGlyLeuGluAlaIleArgLysAlaLeuMetGlyLysTyrAspGlnT 2357
Db 18933 -----CCTATCT 18939
Qy 2357 rpgLugLusProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerL 2377
Db 18940 GCGCGCGCTCGCG-----AGCACCGCGC 18963
Qy 2377 euProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerP 2397
Db 18964 CGCCCGCGCGCTGACCGCGCTGTCGTGTCCTCGCGCATCTGGCGCGACGACCTCCAGCT 19023
Qy 2397 rogly-----GlyGlyGlyLysAlaLysValSerGlyArgProSerS 2411
Db 19024 CGGCCGAGTGCACCGCGAGATGATCGCGCGCGCGGTGAGTTCATGACCGCGCGAGCT 19083
Qy 2411 erArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerVals 2431
Db 19084 CGCGCTGAGCGCGCTCAAGCGCGCGCTCGACGAGC-----AGCAGGT 19128
Qy 2431 erSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpG 2451
Db 19129 CATCGCGGTGCGCGAGCTCGACTGGGAGAGGTACACCGCGCTCTACACCTCGCGCGCGCC 19188
|||

QY 2451 luaspArg-----ProSerSerAlaGlySerThrProPheProTyrAsnP 2466
Db 19189 CACCCCGCTGTTGACAGAGTCCCGAGGTGAGCGGCTACCGCGCTGCGAGCAGA- 19247
QY 2466 roleuileMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProA 2486
Db 19248 -----GGCCCGGGGACCCCGC 19263
QY 2486 laGlySerGlyProLeuAlaGlyProHisAlaTrp----- 2498
Db 19264 CCGGGGCG-----AGTTCCCGCCCGCTGCTGGCCCTGCCCGCCGCAACAGCA 19314
QY 2499 -----AspGluGluProLysProLeuLeuCySer 2508
Db 19315 CCGCAAGCTGCTGAGACCGTCGCGACCGAGCGCGCTCGTGCTCG 19361

RESULT 97
AX703543
LOCUS AX703543 113193 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 1 from Patent WO02097082.
ACCESSION AX703543
VERSION AX703543.1 GI:29538472
KEYWORDS
SOURCE Streptomyces nodosus
ORGANISM Streptomyces nodosus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE
1
AUTHORS Caffrey, J. P.
TITLE Engineered biosynthesis of novel polyenes
JOURNAL Patent: WO 02097082-A 1 05-DEC-2002;
UNIVERSITY COLLEGE DUBLIN (IE)
FEATURES
source
1. 113193
/organism="Streptomyces nodosus"
/mol_type="unassigned DNA"
/db_xref="taxon:40318"

ORIGIN

Alignment Scores:
Pred. No.: 0.000188 Length: 113193
Score: 654.50 Matches: 644
Percent Similarity: 30.30% Conservative: 283
Best Local Similarity: 21.05% Mismatches: 1086
Query Match: 4.95% Indels: 1060
DB: 6 Gaps: 140

US-09-522-753-5 (1-2517) x AX703543 (1-113193)

QY 39 LeuLeuGluTyrGlnHisHisSerArgAspTyr-----AlaSerHis--- 52
Db 11581 CTCCTTCGGCATCAGCGGCACCAATGTCACCGCTCTGGAGCAGCGCGCGGCACAC 11640
QY 53 -----LeuSerProGlySerIleGln 60
Db 11641 CGTGGCCCGCCGCGCCCGCGGCGCACCGCGCGCGTCCCGTGTGTCGG 11700
QY 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
Db 11701 CCGCACCGCGGAGCCCTGG-----CGCCAGCGCGCGCGCTGCT--- 11742
QY 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
Db 11743 -----CACCCACCTCCAGAACCAACCCCGAGCGCTCCCT--- 11775
QY 101 GluPheIleGluSerLysArgProArgLeuLeuLeuProAspProLeuLeuArgPro 120
Db 11776 -----CGCCGACCTCG-----CCACTCCCTGGCCACCCCG 11808
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
Db 11809 CTCCTTCGAGCGCGCGCGCG-----GGTGATCGGCA 11844

QY 141 LeuThrGlyLys---LeuGluProValSerProProSerProPro-----HisThrAsp 157
Db 11845 GG-ACCGCAAGAGACTGCTCGCTCCCTCGGTCTCCTCGCGCGCGCCGACCCGG 11903
QY 158 ProGluLeuGluLeuValProProArgLeu-----SerLys 169
Db 11904 CGGTGCTGAGGTGAGCGCGCGCGGTGCGGTGATGTTCTCGGCGCAG 11963
QY 170 -GluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGlnG 189
Db 11964 GAGCGAGCGCGCGCGCATGGCGCGCAACTGTACGAGACCCAGCCCGGTTCGCGCGCG 12023
QY 189 nileSerLys-----Le 193
Db 12024 CGTTCGAGAGTGTGCGCGCGCTCGACCCGCTGCTCGACCGCGCGTGGCGAGGTG 12083
QY 193 ubLysLysGlnGlnGlnLeuGluGluAlaLysPro----- 207
Db 12084 TGTTCGCGCGCGAGGCGAGAGAGCGCGCTGCTCGACCGCACCGGTGACCACG 12143
QY 208 -----ProGluProGluLysProValSerProProPro 219
Db 12144 CCGCGCTGTCGCGGTGAGGTGCGCGCTCTACGACTTCGAGACTGGGGGTCCGCG 12203
QY 219 eGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaG 239
Db 12204 CCGACTTCGTACCGGTCACTC-----CATCGGTGAGATCGCG 12242
QY 239 uAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnG 259
Db 12243 CCGCGCACATCG-----CGGGGTGTTTCCACTCCAGGACCGCGCGGTGTCGCG 12296
QY 259 nProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLys 279
Db 12297 CCG-----CGCCACCTCATGGAGCGCGTCCGTC----- 12327
QY 279 sLeuIleLeuTyrPheLysArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCy 299
Db 12328 -----CGCGCGCGCATGTGCGCGTGGAGGCCACCGA----- 12360
QY 299 sGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsn 319
Db 12361 -----GGAAG 12365
QY 319 nProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProG 339
Db 12366 AGTCCGCGCGCTGCTCGCGGGGCTGTCGTCG----- 12402
QY 339 uIleArgLysGlnArg---GluLeuGlnGluArgMetGlnSerArgValGlyGlnArgG 358
Db 12403 -----CGCGCTCAACGCGCCACCTCGTGTGTCGTCGCGGACGAGGACCCGCGTGG 12458
QY 358 ySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleLeuaspG 378
Db 12459 AGTCCGCGCGCGAGTTCTCGCGCGGCGCGCGCACCAAGCTGTACCG--GTGAGCCA 12516
QY 378 yLeuSerGluGlnGlnAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProPme 398
Db 12517 TGCCTTCCACTCGCCGACATGAGCCGCTGCTCGAGCGGTTCGCGCGGTTCGCGGAC 12576
QY 398 tLeuTyrAspAlaAspGlnGlnArgIleLysPheIle---AsnMetAsnGlyLeuMetAl 417
Db 12577 GCTGTCTTAGCC--GCCCGCGCATCCGCTGGTCTCGACCTGACGCGCGCGGCGG 12633
QY 417 aAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGlu 437
Db 12634 CGACGACGCGGAGGTG-----CGCACCGCGCTACTGCTGGTGGCGCCGCTCCG 12681
QY 437 sGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPh 457
Db 12682 TGAGCGCGTACCG----- 12694

Db 14470 CCCCCGCGC-----CCGCCCTCACGAGATGCGCGCGCGTGGTCCGCCAGGACGT 14520
Qy 1069 laProaspProSerAlaPheSer-----TyAlaProProGlyH 1082
Db 14521 CACCGTACCGTCCGGAGAGTCACTGGACCGCGCTAGCGCGCGCTGTTCACCTCCGCCG 14580
Qy 1082 isProLeuProLeuGlyLeuHis---AspThrAlaArgProValLeuProArg-----P 1099
Db 14581 GCCAGCGCCCTCATCAGCGACCTGCCGAGGTCCGTCGCTGGCGCGCGAGAACACCC 14640
Qy 1099 roProThrIleSerAsnProProProLeuIleSerSerAlaIysHisProSerValLeuG 1119
Db 14641 CGCCGACCGCGCGACCGCTCCGAG-ATCGTCCAGCGGGTCCGCTCGCTGTCCGAGCCG 14699
Qy 1119 luArg----- 1120
Db 14700 AGCAACTGGCGCTGCTCACGAGCTGGTGGCAACGAGGCGCGGACCGTGTCTCGTCACT 14759
Qy 1121 --GlnIleGlyAlaIleSerGln-----GlyMetSerValGlnL 1133
Db 14760 CCTCCCGCGCGCGTACCGGAGGACCGTGCCTTCCCGGAGATCGGCTTCGACTCGCTGA 14819
Qy 1133 euHisValProTyrSerGluHisAlaIysAlaProValGly-----ProValThrM 1150
Db 14820 CCGCGCTCGAACTCGCGCAAGCACTCGCGCGCGCCACCGGCTCTCGCTGCCGAGACCA 14879
Qy 1150 etGlyLeuProLeuProMetAspProIysIysLeuAlaProPheSerGlyValIysGlnG 1170
Db 14880 T-GGTGTTCGACTACCC--GACCCGCTGGAGCTGGCCCGCAGTATCTGGGGCCGAGATG 14935
Qy 1170 luGlnLeuSerProArgGlyGlnAlaGlyPro----- 1180
Db 14936 GTCGGCTCGCTCTGGAGGTGCG--GGGCCCGCTGCGCACCGCGGTACCGACGACGAA 14992
Qy 1180 ----- 1180
Db 14993 CCCATCGCCATCATCGGCATGAGTCCGCTACCCCGCGGAGTCACTCCCGGAGCAG 15052
Qy 1181 -----ProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyT 1198
Db 15053 CTGTGGACCTGTGTCTCCG-GCACCGAGCGGATCACCAGATTCC-----CGGTCAA 15105
Qy 1198 hrAlaLeuGlySerValProGly----- 1205
Db 15106 CCGCGGTGGAAACACCGCGGGCTTACGACCCCGACCGACCCCTGGCCACCACTTA 15165
Qy 1206 -----GlySerIleThrIysGlyIleProSerThrArgValProSerAspS 1221
Db 15166 CTCACGCGAGGCGGCTTCTGACAGAGCGCGAGATTGACCCCATGTTCTTCGGCAT 15225
Qy 1221 erAlaIleThrTyrArgGlySerIleThrHisGly----- 1232
Db 15226 CTGCGCCCGGAGCGGTGTGTATGACCGCGAGCGCGCTCTCTCTGGAGACCACTG 15285
Qy 1233 -----ThrProAlaAspValLeuTyrIysGlyThrIleThrA 1245
Db 15286 GGAGGCTTGAACGCGCGGACTCACCCCG-GACACCCCTC---CGCGGACGCTGACCG 15341
Qy 1245 rgIleIleGlyGluAspSerProSerArgIleAspArgGlyArgGluAspSerLeuProL 1265
Db 15342 GCACCTTCATCGGCTCCAGCTACCGAGTACCGGATGGCGCGCGGACGCGC---GCCG 15398
Qy 1265 ysGlyHisVal-----IleTyrGluGlyLysLysGlyHisValL 1278
Db 15399 AGGGCCACCTGGTCAACCGGCACACGCCCGAGCTGTCTCCGCGCGCTCGCTATGTCT 15458
Qy 1278 euSerTyrGluGly---GlyMetSerVal---ThrGlnCysSerLysGluAspGlyArgS 1296
Db 15459 TCGGCTGGAGGCGCGCGGTACCGCTCGACACCGGCTCTGCTCTCGCT--GGTCGCC 15517
Qy 1296 erSerSerGlyProProHisGluThrAlaIaProIysArgThrTyrAspMetMetGluG 1316
Db 15518 CTGACCTGCGCTGCCA-----GGCGCTGCGCAACCGCGCGGACGACCACTTCGCGGTGCGG 15571

Qy 1316 lyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleP 1336
Db 15572 GCGCGCGCACCGTCATGACCAACCCCAACGCGTTCGTGGCTGACGCGCGAGCGCGC 15631
Qy 1336 ro-----ProGluArgHisSerProHisH 1344
Db 15632 CTCGCGCAGCAGCGCGGTGCAAGCGTTCTCCGAGAGCGCGCATGACCTCGCC 15691
Qy 1344 isLeuIysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrV 1364
Db 15692 GA---AGGTGTCCGATCTCTCTGTCGAGCGGCTGTCGAGCGCGCGCGAACGCGCAC 15748
Qy 1364 alGluAlaGlnGluAsp---TyrLeuArgArgGlu---AlaIysLeuLeuIysArgGluG 1382
Db 15749 CCGGTGCTCGCGTGTATCGCGGCTCCGCATCAACACGAGCGCGCTCCCAACGCGCTC 15808
Qy 1382 lyThrPro-----ProProProPro----- 1388
Db 15809 AGCGCCCCCAACGCGCCCTCCACGAGCGCGTCATCCGCGAGCGCTTGCCCAACGCGCG 15868
Qy 1389 -----ProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysL 1407
Db 15869 GTCGCCCCCGCGGAGATCGACCTCTCGGA-----GGCCCCACGCGACC 15910
Qy 1407 euIysProAla-----HisGluGlyLeuValAlaThrValIysGluAlaGly--- 1422
Db 15911 GGCACCCGCTCGCGGACCCCATCGAGG---GCAGGCGCTG 15949
Qy 1423 --ArgSerIle-----HisGluIleProArgGlu-----GluL 1433
Db 15950 TTGCGCACCTACGCGCGCACCGGTACCCCGGAGACCGCGCTGCTCGCTCGGTGAA 16009
Qy 1433 euArgHisThrProGluLeuProLeu----- 1441
Db 16010 TGAACATCGG--GCACCTCCAGTCCGCGCGGTGTCGCGAGCATCATCAAGATGGTC 16066
Qy 1442 -----AlaProArgProLeuIysGluGlySerIleThrGlnG 1454
Db 16067 ATGGCGCTCGCCACGCGTGTGATGCCCGACACCTGCACGCGAGCGCTC----- 16119
Qy 1454 lyThrProLeuIysTyrAspThrGlyAlaSerThrThrGlySerIysHisAspVal- 1473
Db 16120 --CAGCCACGTCTGACTGGAGCGCGGACCGCTACGCGTCTCTCGCGAGAACACGACTGG 16177
Qy 1474 -----ArgSerLeuIleGlySerProGlyArgThrPheP 1485
Db 16178 CCGCAGACGCGCGTCCGCGCGCGCGGTGCTCTCTCGGCATCAGCGTACCAAC 16237
Qy 1485 roProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrG 1505
Db 16238 GCCACGTCTCT-----GGAGCAGGAGACCGAGCGCGCGCC 16276
Qy 1505 luGluSerLeuIysSerArgProGlyThrAlaSerSerGly-----GlySerIleA 1523
Db 16277 GCGAGGACGAGCAGCTCGCGCGCGCGCTGCGCGTCCGCGCGCGGTGTGGTGGCTGG 16336
Qy 1523 laArgGlyAlaPro-----ValIleValProGluLeuGlyLysProArgGlnSerProL 1541
Db 16337 CTGCTCTCGCGCTGGCGCGCGGTGCGCGAGCAGCGCGCTCTCTCCTCCTCACCCT 16396
Qy 1541 euThrTyrGluAspHisGlyAlaPro-----PheAlaGlyHisLeuP 1555
Db 16397 CTGTCACCGCGCACCGCGCGCGCTCGACATCGGTCTGCTCTCTCGCGCACCTCC 16456
Qy 1555 roArgGlySer-----ProValThrMeta 1563
Db 16457 CCGCGCGC-TCTTCGACACCGCGCGCTGTGTGTCGCGCGCGCGCGCACCGCGCTGGA 15515
Qy 1563 rgGluProThrProArgLeuGlnGluGlySerLeuSerSerLysAlaSerGlnAspA 1583
Db 15516 GCGCTGCGCGCGCTCGCGCGACCGGCGCTCGCGCGCG----- 15655

Qy	1583	rgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValProGluH	1603
Db	16566	-----GTGTCGGACGTCGGGCGCGACCGTCTTCTCCCGGCA	16611
Qy	1603	iHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspL	1623
Db	16612	GGGCTCCC-----AGTGG-----	16625
Qy	1623	eutYrArgSerHisIleProLeuAlaPhe---AspProThrSerIleProArgGlyIleP	1642
Db	16626	-----CGGGATGGCGCCAGCTCTCGACGAGTCCCGGTGTTCGCGCGAGCGATCGC	16680
Qy	1642	roLeuAspAlaAlaAlaTyTyLeuProArgHisLeuAlaProAsnProThr-Tyr	1661
Db	16681	CG-----AATGCGCGCGCGCCCTCGCGGAGTTCACCGACTGGAA	16719
Qy	1662	ProHisLeuTyProProTyTyLeuIleArgGlyTyTyProAspThrAlaAlaLeuGluAsn	1681
Db	16720	CCTCATCGAGCTCTCGCGGGCGC---CGAGGCGCGCCGACGCTGGAGCGCGTCGACGT	16776
Qy	1682	Arg-----GlnThrIleIleAsnAspTyTyIleThrSerGlnGlnMetHisAsnThr	1699
Db	16777	CGTCCAGCCGCTCGTTCGCTCGATGTCTCCCTGGCGCGCGTGTG---CGGGGCCCA	16833
Qy	1700	AlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSer	1719
Db	16834	GGGTGTGCAACCCGACGCGGTGTGCGCCACTCGCAGGCGGAGATCGCGCGCGTGTGT	16893
Qy	1720	LeuAlaLeuAsnTyTyAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHis	1739
Db	16894	CTCCGCGCCCTGTGCTGCGGACGCGCCCGGTCGTGACGCTCCGCGCCAGCCCAT	16953
Qy	1740	LeuProValLeuValPro-----ProThrPro---1748	
Db	16954	CGGCGCA-CCCTCGCGCGCGCGCGGATGATGCTGTCGGCTGCCCTGCGCGGAG	17012
Qy	1749	-----GlyThrProAlaThrAlaMetAspArgLeu	1758
Db	17013	TGGAGGCGCTGTGGAGCGGTTCGAGGAGCGGTGTCTCGTCGCGCGGAGA-----	17063
Qy	1759	AlaTyTyLeuProThrAlaProGlnProPheSerSerArgHisSer-----	1773
Db	17064	-----ACGACCCCGCTCTCTCGCTCGTCGCGGCGAGCCGAGCCCTGTGACG	17111
Qy	1774	-----SerSerProLeuSerProGlyGlyPro-----Thr	1783
Db	17112	AATGTCACGCACAGCTCACCGCGAGAGATCGGGCCCGCGGGTGCCTGTGACTAG	17171
Qy	1784	HisLeuThrTyPro-----ThrThrSerSerSerGluArgGluArg	1798
Db	17172	CCTCGCACTCCCCCATGTGAGGACCTGCAGCAGAGATCTCGAACTGTCTGGCCGAG	17231
Qy	1799	AspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThr	1818
Db	17232	TCGCGC-----CGCGACGCTCCGAGATCCCGTCTCTCTCGACCGTCACCG	17276
Qy	1819	ThrThrValGluHisAlaProIleTyTyArgProGlyThrGluGlnSerSerGlySerSer	1838
Db	17277	GCAGCTGGCTGACACACCGTCATGAGCGCGGTACT-----	17315
Qy	1839	GlySerSerGlyGlyGlyGlySerSerSerArg-----	1850
Db	17316	GGTACCGCAGCTGCGCGCGGTGCTGTTCGCGGACGACGTCGCTGACCTGATCGCG	17375
Qy	1851	-----ProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArg	1866
Db	17376	CCGACCACCGTGCCTTCATCGAGTCACTGCCACCCCGTGTCTCGCCATGTCCGTCCAG	17435
Qy	1867	ThrGlnAspAlaLeuGlnGln-----ArgProSerValLeuHisAsnThrGly	1882
Db	17436	ACATGATCGACGAGCGCGGAGTGGCGGTGTGCTCCGCGACCTCGCGCGCAACAG	17495
Qy	1883	MetLysGlyIleIleThrAlaValGluProSerLys-----ProThrValLeuArgSer	1900

Db	17496	GCGGTCTCGACCGCTTCCTGCTCTCCGCCCGCAGAGGTTCCTGCTCCGCGTGTCACAGTCG	17555
Qy	1901	ThrSer-----ThrSerProValArgProAlaAthrPheProAlaThr	1917
Db	17556	ACTGGGCGCGGTGTTGAGGGACCGGTGCTCCCGGTGCAGCTGCCCACT	17609
Qy	1918	HIScysProLeuGlyGlyThrLeuAspGlyValTyPrProThrLeuMetGluProValLeu	1937
Db	17610	--ATGCTTCCAGCACGAGAACCCTGTGGCCCATGGCCGCCCGCGAGGCG--	17666
Qy	1938	LeuProLysGluAlaProArgValAlaAArgProGluArgProArgAlaAsthRgLyHis	1957
Db	17661	TCACCGCGCGCACCCGAGGACGCGCT	17690
Qy	1958	AlaPheLeuAlaLysProProAlaAArgSerGlyLeuGluPro-----Alaser	1973
Db	17691	-----TCTGACCGCGCTCGAGAGCGGAGACGTCTCCGCCCTGACCGCGGCCTCG	17741
Qy	1974	SerSerLysGlySerGluProArg-----ProLeuValProProValSerGlyHis	1991
Db	17742	GACCGACGAGGACTCGCTCGCGCGCTGCTGCCCGCTCTCTCTGCGCGCGGCC	17801
Qy	1992	AlaThrIleAla-----ArgThrProAlaLysAsnLeuAlaProHisHisAlaserPro	2009
Db	17802	GCAAGGAGCGCTCCACCGTCACTCTGCGCTTACCGCCGACCTGGAAGCCCGTCACCA	17861
Qy	2010	AspProProAlaProProAlaserAla-----SerAspProHisArgGluLys	2025
Db	17862	AGTCCCGCAGCGACCTCGGACGGACCTGGCTGCTGCTGCTGCCCGACGGGCTGACG	17921
Qy	2026	ThrGlnSerLysProPheSerIlegInGluLeuGluLeuArgSerLeuGlyTyHisGly	2045
Db	17922	ACCGACGTCCGCGAGGCGCTGAGACGGCGCGCGCGCTGCGCC-----GSC	17972
Qy	2046	SerSerTy-----	2048
Db	17973	TCGTCTTGACGAGAGCTGCACGACCGCGCGCTGCTGCGGAGCGCCTCACCGAGCGCG	18032
Qy	2049	-----SerProGluGlyValGluProValSerProValSerPro	2062
Db	18033	ACGGCTGACCGGATGCTCTCGCTCTCGGGCGCGGAGCGGACCGGTGCGTCCCCG	18092
Qy	2063	SerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSer-----	2080
Db	18093	GCACCGGGCTCGTCTCGGGTGCCTCACGCTCGCCCTCGTCCAGGGCGCTCGCGACG	18152
Qy	2080	-----	2080
Db	18153	CCGATAATCGACACCCCGCTGTGGCGCTGACCCGCGCGCGCTCTCCACCGCGCGCTCG	18212
Qy	2081	--HisLeuGluGlyGluLeuArgProLysGlnPro-----GlyProVallys---	2095
Db	18213	ACAAGTCAACCGCGCGCTCCAGCGCCAGGTACCGGATCGGCTGACCGCGCGCTCG	18272
Qy	2096	-----LeuGlyGlyGluAlaAlaHisLeuPro---HisLeuArgProLeu---	2109
Db	18273	AGTCCCGGACGCTGGGGCGGTGCTGCTGACCTGCCCGGACACTGGACGCCCGCGCG	18332
Qy	2110	-----ProGlu	2111
Db	18333	GCCAGCGCTCGCGGCGCTCTCGCGCGCGCGCTCGGGCGACGACGACCATGCGCCTGC	18392
Qy	2112	SerGlnProSerSerProLeuLeuGlnThrAlapProGlyVallysGlyHisGlnArg	2131
Db	18393	GCTCTCCGCGCTTCACCCGCGGTATCGTCCGCGCG--	18431
Qy	2132	ValValThr-LeuAlaGlnHisIleSerGluValIleThrGlnAspTyThrArgHisHi	2151
Db	18432	-----ACGCGCGCGCGGCTCGCGCGGACTG-----GAAGCGCGCGGACCA	18479
Qy	2151	sProGlnGlnLeuSerAlaProLeuProAlaPro-----	2162

```
Db 18480 CCCTGCTCACC - GCGGCTCGGCGACCCCTCGCCCCGACCTCGCGCGCTGGCTCGCCGAG 18538
Qy 2163 -----LeuTyr-SerPheProGlyAlaSerCysProValLeuAspLeuArgAr 2178
Db 18539 CAGGGCGCCGAACACCTGCTCTGCTGTCACCGCGCGCGGCCG-----AGGCG 18586
Qy 2178 gProProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerPr 2198
Db 18587 CCCGGCGCCGCAACTCCGCGCGAGCTGGCGGAACGCGGACACGAGCAGCCTCGCC 18646
Qy 2198 oHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerVal-LeuGlyGlyG 2218
Db 18647 GCCTGCGACATCACCGACCGCTGACCGCTCGCGCGCTCTCGAAGCCTGAAGCCGAG 18706
Qy 2218 lylGluAspGlylleGluProValSerProProGluGlyMetThrGluProGlyHisSerA 2238
Db 18707 GCGCG-----CACCGTCGCGACCGTGGTGACACCGCGCGCCACCATC 18748
Qy 2238 xgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetG 2258
Db 18749 GAACCTGCA-----CACCTCGAGGCCA-C 18771
Qy 2258 lySerLysSerProGlyAsnThrSerGlnProProAlaPheSerLysLeuThrGluS 2278
Db 18772 CACCTCGACGACTTCGACCGCTCTCGCGCGCAAGGTCACCGCGCGCAGATCTCGA 18831
Qy 2278 erAsnSerAlaMetValLysSerLysGlnGluileAsnLysLysLeuAsnThrHieA 2298
Db 18832 CGAACCTCTCGACGACGAGGAGC-----TGACGACACTTCGT 18867
Qy 2298 snArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProA 2318
Db 18868 CCTGACTCTCTCA-----CCGC 18885
Qy 2318 laIleThrGlyThrGlyLeuMetThrTyrArg---SerGlnAlaValGlnGluHisAlas 2337
Db 18886 CGGCATCGCGGACGCGCGCACCGCGGTATGTGCGCGGCAACG----- 18932
Qy 2337 erThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnT 2357
Db 18933 -----CCTATCT 18939
Qy 2357 rpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerL 2377
Db 18940 GCGCGCCTCGCG-----AGCACCGCGC 18963
Qy 2377 euProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerP 2397
Db 18964 CGCCCGCGGCTGACCGCCTGTGCTGCTGCGGCATCTGGCGCGACACCTCCAGCT 19023
Qy 2397 roGly-----GlyGlyGlyLysAlaLysValSerGlyArgProSerS 2411
Db 19024 CGGCCGAGTCGACCCGAGATGATCCCGCAGCGGCTCGACGACGAC-----AGCAGGT 19083
Qy 2411 erArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValS 2431
Db 19084 CGCGCTGAGCGGCTCAAGCGGCGCTCGACGACG-----AGCAGGT 19128
Qy 2431 erSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpG 2451
Db 19129 CATCGCGTCCCGACGCTCGACTGGGAGAGTACACCCCGCTCTACACCTCGCGCGCGGCC 19188
Qy 2451 luAspArg-----ProSerSerAlaGlySerThrProPheProTyrAsnP 2466
Db 19189 CACCCGCTGTTCCAGCAGGTCCTCCGAGGTGACGCGCTCACCGCGCTCGCGAGCAGA- 19247
Qy 2466 roLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuProA 2486
Db 19248 -----GCGCGCGCGGACCCCGC 19263
Qy 2486 laGlySerGlyProLeuAlaGlyProHisAlaTtp----- 2498
Db 19264 CCGGCGCG-----AGTTCCGCGCGCGCTGCTGCGCTGCGCGCGCGCGGACGACA 19314
```

```
Qy 2499 -----AspGluGluProLysProLeuLeuLeuCysSer 2508
Db 19315 CCGCAAGCTGCTGGAGACCGCTCGCACCGAGCGCGCTCTCGTCTCG 19361

RESULT 98
SPSNBCDE 22449 bp DNA linear BCT 25-APR-1997
LOCUS S.pristinaespiralis snbc gene & snbDE gene.
DEFINITION Y11548.1 GI:2052247
ACCESSION Y11548.1
VERSION Pristinamycin I synthase 2; Pristinamycin I synthase 3 and 4; snbC
KEYWORDS gene; snbDE gene.
SOURCE Streptomyces pristinaespiralis
ORGANISM Streptomyces pristinaespiralis
Bacteria; Actinobacteridae; Actinobacteriales; Streptomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS de Crecy-Lagard, V.A., Saurin, W., Thibaut, D., Gil, P., Naudin, L.,
Crouzet, J. and Blanc, V.
TITLE Streptogramin B biosynthesis in Streptomyces pristinaespiralis and
Streptomyces virginiae: molecular characterization of the last
structural peptide synthase gene
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 22449)
AUTHORS de Crecy-Lagard, V.A.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-1997) V.A. de Crecy-Lagard, Institut Pasteur,
Unit de Biochimie Microbienne, 28 rue du Dr Roux, 75724 Paris
cedex, FRANCE

FEATURES
source
1. .22449
/organism="Streptomyces pristinaespiralis"
/mol_type="genomic DNA"
/strain="SP92"
/db_xref="taxon:38300"
/clone_lib="PIBV3 cosmid"
115. .7890
/gene="snbC"
115. .7890
/gene="snbC"
/codon_start=1
/transl_table=11
/product="Pristinamycin I synthase 2"
/protein_id="CAA72311.1"
/db_xref="GI:2052248"
/db_xref="GOA:O54959"
/db_xref="SPTREMBL:Q54959"
/translation="MLVDDARLPLTAQLGTFWFAQRDPHSPVNTGEYVELTGLD
ADRFPAALRATVARTDALRLRPHRTDAEGTIVTSVADDVFTLHRVDVSDEDDPAAAD
AMWRDVIARPLDVTCTGRLFTQALFTLGFGRHWTFQHVHLLLDGFGFLVAERVAAD
VRAGAEAGDAEAGDTEPGIRDLIAAEADWRSDAYAQDRAHWLATPRHTPTVLGTTA
PCAPALRRTASAGPEVLADVAARAEATGQVAFPAAVAAVVOVLTGERTVTLGMP
VTARTISAARRAPGTVANVPALHVSERPAAPLAEIVASVRDALEPLLRHVPYEBAR
IDLEANAAYGESETDAHLARVFLGAFEAAGPDPTPLGRHAAPAAAPAAAPPTGG
PALHTLFAARAAATPHAVASGLDGDHLTYAQDLARANLRLHVEGAGPERLVALTL
PSRLDLVALLAVLKSAAVPLDPTPADRLAMLEDAAPVLVLDPOQAMTDLGYP
DTAPRTAVDGHPAVVIYTSGSTGRPKGVVIPSNNVRLFTSTDHWFSGPDDVWTLF
HSVADPFWMEIHWGALLHGGRLVVPVHVSRSFGPDLDLAREKVTVMQPTAFHQL
DAADRARTAAPEALARYVFGGEALDVAELADYARRGTAARLVNMGITETTVHVT
ALPGHGVPTGSLGGGIPDLTPRVLDAALRPVPGFTGELYVYAGAGLASYLNRH
ALTQFPADPYGAPGRMYRTGDLVRLDDGTAYLGRGDQDQVKIRGRFRIELGEIEN
VLATHEGQAQAANVRDRHGDRLRLAAYVPTGTPGVAGRWALLAGOLUPHSLPAS
FTLLAALPLTAKGLDRAALPAQAPAPETADEPADPGEAALRRI FADTLGTFRV
GADSDFTLGGHSLLAGRLIAAVRDTLGAETIQDLFEAPTAGLAARVTRATGTRPP
LVPRHGDGPRPTSVAQHEMWVLREVEGAAPTYNMPLEALRLSCTLNRTLRAALYDV
VRHESLRTVYPRGEGTVLRLLPPEQARPVWIESRVDDQLPAALTAAREEPNLDCQ
LPIRAHLFTLPGGDSVLLIMHHVAGDWSLRPLGLDLSIAAARLAARAPPPPLPV
SVADFAVWQRELIGDEHPSLAARQALHCHWLQAAALLTRLGAGTDTPI CGCTASRPDS
SVPLAVPPQALNLDTRFAHAGTACHWLQAAALLTRLGAGTDTPI CGCTASRPDS
ADLDLVGFVNPWLRTADTSGDPTFTELLARATALAAAHQDVFDFVRLEQPER
SAAHPLFLQVLLSYQDFEARHLKGLDVRREDVDFGVAKFDLTFLNLVHRHGDEEPAG
```


QY 88 SerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArg 107
Db 12902 ACTCGCGGACCGTCCCGCCAGCGCCGCGCTTCGAGTGGGACCCACAGCGCGCT 12961
QY 108 ProArgLeuGluLeu----- 112
Db 12962 CTTGAATCTCGCTCCATAGCTTCAGCGCCGCGTGTGGAGCTGACCTCCGCGTGT 13021
QY 113 -----LeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGly 127
Db 13022 CACCGCGCGGAGTCTGCTCTGCGCAGCGCGAGCTGCTGCGCGCGCGTGTGGC 13081
QY 128 GlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluPro 147
Db 13082 CCGCTCTGTCACGACCGGCACATCACCCTGATCGC-----CCTCGCGCGCTTCGCGCT 13135
QY 148 ValSerPro----- 150
Db 13136 GCCCGCTCTCCGACGCGGCCCTGCGCGCGGAGCGGACCTCATCTGTCGCGGCGACGC 13195
QY 151 -----ProSerProProHis----- 155
Db 13196 CACGCGCGCGACGCGCGCTTTCGACACCGCGCGCGCATGTGTCAACGCTACGG 13255
QY 156 ThrAspProGluLeuLeu----- 162
Db 13256 ACTGACGAGACACCGCTGTGCGCCACCATGAGCGAGCGCGCCACCGGCGCGCGCC 13315
QY 163 -----ValProProArgLeuSerLysGluLeuIleGlnAsnMetAspArgValAsp 180
Db 13316 GCCATCGCGCGCGCTGCCACGCGCGCGCTACGTCTCGGACGACGCGTGGCGCC 13375
QY 181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnLeu 200
Db 13376 CGTACCGCGCGGTACCGCGGAGATGACGTCTCGCGCGCGGTCTCGCGCGCTA 13435
QY 201 GluGluAlaAlaLysProGluProGlu-----LysProVal 214
Db 13436 -----CCTGACCGCGCGCGCTGCGCGCGCGCGCTGTCGCGCGCGCGCTA 13483
QY 215 SerProProIleGluSerLysHisArgSerLeuVal---GlnIleIleTyrAspGlu 233
Db 13484 CGCCTCTGTTTCGCGGACCGCGCACCGCATGTACCGACCGCGACCTGGCAGCGCT 13543
QY 234 AsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGlu 253
Db 13544 GGACGCGGACGCGCGGTGCACTTCGCGCGCGCGCGCGCGCATCAAGATCCGCGG 13603
QY 254 LeuPro-----LeuTyrAsnGlnProSerAspThrArgGln 265
Db 13604 CTTCCGATGAAACCGCGGAGATCGAGACGCTCTGACCGCGCACCGCGCGTGGCGG 13663
QY 266 TyrHisGluAsnIleLysIleAsnGlnAlaMetArg-LysLysLeuIleLeuTyrPheLy 285
Db 13664 CGGCGCGTGTATCCCGCGGAGGACACCGCGGCGCAACAGCTGTGCTCTACCTGAC 13723
QY 285 sArg-----ArgAsnHisAlaArg-----LysGln 293
Db 13724 CGCGGACACCAACCGCGCGCGCGCGGACCGGACCGCGCGCGCGCGCGCGCGG 13783
QY 293 nTTPysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysVal 313
Db 13784 CTGGCAG-----CAGACCTAGATCTCTCTAGAGCGCGCAACCGCGCGCGCGG 13831
QY 313 1-----GluArgIleGluAsn---AsnProArgArgArgAlaLysGluSerLysValAr 330
Db 13832 CTTCCGCGGAGACTTCTCGCGCTGMAACAGCAGCTACACCGCGGAGGATCCCGCTGGA 13891
QY 330 gLutTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnArgMe 350
Db 13892 GGAGATGCGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 13951

QY 350 tClnSerArgValGly----- 355
Db 13952 GGTCTTGAGATCGCTGGGACCGCGCTGATCTGTCCAGTTCGCCGCCACACCGA 14011
QY 356 ---GlnArgGlySerGlyLeuSerMetSerAla----- 365
Db 14012 GGAGTACCGGGGACCGGACTCTCCCGAGCGTCTGTCGCCAGGCTGGCGGCCACCTCGC 14071
QY 366 -----AlaArgSerGluHisGluValSe 373
Db 14072 CGCGCGCGCGACCTGGCGGACAGGTCACTGAGACCGCGCGCGCGCACGACGACGA 14131
QY 373 rGluIleIleAspGly-----LeuSerGluGlnGluAsnLeuGluLy 387
Db 14132 CGACCTGCCAAGGCGCTTCGACACGATCTCTCAACTCCGTCGCGCAGTACTTCCC 14191
QY 387 sGlnMetArgGlnLeuAla-----ValIleProPro----- 397
Db 14192 CGACGCGCGTACTCTCGCGGATCTCCACCGCGCGCGCGAATCTGTGGCGCGCGGG 14251
QY 398 -MetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsn----- 411
Db 14252 CACGATCTTCTCGCGCATCTCCGAACCTGGCGACCTGTGACGCTTCGCGACCGCGT 14311
QY 412 -----MetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnVa 428
Db 14312 CGAATCGCGCGCGCGCGCTTCCGCGACCGCGCGCGTA-----CGCGCGC 14362
QY 428 lMetAsnMetTyrPsrGluGlnGluLysGluThr-----PheArgGln 442
Db 14363 CDTCAACAGTCACTGCGCACCGGAGAACTCTGTCTGACCGCGACTTCTTCCCGA 14422
QY 442 uLysPheMetGlnHisProLysAsnPhedGlyLeuIleAlaSerPheLeuGluArgLysTh 462
Db 14423 CTTCTGCGACCGCGCGCGCGC-----CTGACCGCGCGCATCATCTCGCGCAC 14473
QY 462 rValAlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLe 482
Db 14474 GGCGGCGCAC-----CACAAACGAGATGAGCGCGCACCG 14506
QY 482 uValArgArg----- 485
Db 14507 -TACGACGTACCTTCCACCGCGCGCGGACAGCCCGCGCGGAGACGCTCTGC 14565
QY 486 -----SerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGln 500
Db 14566 GCTGGGGGACCGACCTCGACCGCGCGGAACTCGCGCGCTCTCTCGCGCGCGCGC 14625
QY 500 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Db 14626 CGTGGCGCTACCGCGCATCCCGCGCGCTCACCGCGGAGACACGCGCGCACAG 14685
QY 520 sAsp-GluLysGluLysGluLysGlu-----AlaGluLysGluGluLysProGluV 538
Db 14686 CCTCGAACCGGATCGCGCGCGCGCTGCGCGCGCGCTGAGCGCGCGCGCGCGCGG 14745
QY 538 alGluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspA 558
Db 14746 GCGCATCTGACCCCGACGACGAGTGTAGCGTGGCGCGCGCTGTGCACCTGGGGCCAGC 14805
QY 558 snAspGluLysGluAlaValAlaSerLysGlyArg-LysThrAlaAsnSer---GlnGly 576
Db 14806 ACGACACCTTCGACGCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14865
QY 577 ArgArgLysGlyArgIleThrArg-----SerMetAlaAsn 588
Db 14866 AGGACCGCTTCGCGACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14925
QY 589 GluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAla----- 605
Db 14926 TGGCGGACCTCTGTCGCGCGCGCGG-CTGCGCGCGCGCATGCGCGCGCGCGCTACGCTCTG 14984
QY 606 -----SerMetGluLeuAsnGluSerSerArgTyrThrGluGluMetGluThrAla 623

QY 1152 LeuProLeuPro-----MetAspProLysLys 1160
Db 17040 GACCGCTGCCGACGACCTGGGCGCTACCGGACACCGACCTACCGACCGCGACCGC 17099
QY 1161 LeuAlaProPheSerGlyValLysGlnGln-----LeuSerProArgGly 1176
Db 17100 ACGGCGCG-----CTACGGCGGACACCGCGGTACGTATCCACACCTCCGGC 17150
QY 1177 GlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGlnAlaSerValLeuArg 1196
Db 17151 TCCACCGGACACCGGCGGTGCTATGCTCCGCGCGCTGCGCAACCTGCTGACC 17210
QY 1197 GlyThrAlaLeuGlySer-----ValProGlyGlySerIleThrLysGlyLeuProse 1214
Db 17211 TGGCAGCGCGCGCTTCCCGCGCGCACCGGGGTGCGCA-----CCGCC 17255
QY 1214 rThrArgValProSerAspSerAlaIleThrThrArgGlySerIleThrHisGlyThrPr 1234
Db 17256 CAGTTACCGCCATCGGCTTCGACTTCTCGGTGACGAGATCTCTCCCGCTCGTCATG 17315
QY 1234 oAlaAspValLeuTyrlsGlyThrIleThrArgIleGlyGluAspSerProSerAr 1254
Db 17316 GGCA-----AGACCTTCGCC 17330
QY 1254 gLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrlsGlyLysLys 1274
Db 17331 GTGCCCTCGAAGAGTCCGCCACGCGCGAAC-----TGTCGGCGCGTGGCTC 17381
QY 1274 sGlyHisValLeuSerTyrlsGlyGlyMetSerValThrGlnCysSerLys-----G1 1292
Db 17382 GAGACACGACGATCA-----ACGAATGTCGCGCGAACCTC 17420
QY 1292 uAspGlyArgSerSerGlyProProHisGluThrAlaAlaProLysArgThr--Tyr 1311
Db 17421 GTCATCGAGCGCTCGCGAGCGCGCGAGCGCGCGACCTCGCGGACCTCAC 17480
QY 1312 AspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMet 1331
Db 17481 GATCATCTCCAGGG-----GGCAGGCGCTCGCCCCACCGGAGGTACGCGCTTC 17534
QY 1332 GlyArgAlaIleProGluArg---HisSerProHisLysLysGluGlnHisHis 1350
Db 17535 ACGCGCGCTACCGCGCGCGCTCGCAACGTGTACGCGCGCGGAGACCGACGCC 17594
QY 1351 lLeArgGlySerIleThrGlnGlyIleProArgSerTyrlsValGluAlaGlnGluAspTyrl 1370
Db 17595 GTCACACCCACACCTCGCGCGCTGATCCGCGCACCTGG-----17633
QY 1371 LeuArgArgGluAlaLysLeuLeuLysArgGlyThrProProProProProSer 1390
Db 17634 -----CCCCCTCCGACCATCGGC 17654
QY 1391 Arg-----AspLeuThrGluAlaTyrlsThrGlnAlaLeuGlyProLeuLysLys 1408
Db 17655 CGGCCCTGCACGACGCGGTGTAGTCTCTGGACAGC-----GCACCTGGC 17702
QY 1409 ProAlaHisGluGlyLeuAlaAlaThrValLysGluAlaGlyArgSerIle-----1425
Db 17703 CCGCTCCCGCGCGGTCAACCGCGAGCTGTACTGCGCGCGCGCGCGCTCGCGCGCG 17762
QY 1426 -----HisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaPro 1443
Db 17763 TACTGTACCGC-----CCGCGACACACCGCGAGCGGTTCGTGCGGACCTACGCCCCA 17819
QY 1444 ArgPro-----LeuLysGluGlySerIle--ThrGlnGlyThrProLysLys-- 1458
Db 17820 CGGCCCGCGCGCATGTACCGACCGCGGACCTCGCGCGCTGGAACACCGCGCGGAG 17879
QY 1459 -----TyrAspThrGlyAlaSerThrThrGlySerLysHisAspValArgSerLeuIl 1477
Db 17880 CTGGAGTTCCGCGCGCGCGACACCGAGTCAAGATCC-----CGCGCTTCGCC 17930

QY 1477 eGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaAr 1497
Db 17931 GTGGACCGCGGAGTTCGAGCGCGCTTCACCGCCACCGCGCATCACCGAGGAGCG 17990
QY 1497 gAlaLeuGluAlaArgAlaCysTyrlsGluGlySerLeuLysSerArgProGlyThrAlaSerSe 1517
Db 17991 GTCC-----TCGCGACGCGGACCGCTCGTC 18017
QY 1517 rSerGlyGlySerIleAlaArgGlyAla-----ProVa 1528
Db 18018 GCCTAGCTGCTCACCAACCGCGACCGACCTGACCGGGGTACGCGAGCACCTGGCGGCC 18077
QY 1528 lIleValProGluLeuGlyLysPro-----ArgGlnSerProLeuTh 1542
Db 18078 CGCTCCCGAGCTTCATGTCTCCACCGCTAGCTCGGCTCGAGGCGGTG--CGCTCAC 18136
QY 1542 rTyrlsGluAspHis-----GlyAlaProPheAlaGlyHi 1553
Db 18137 CGTGAACGGCAAACTCGACCGGCGCGCTTCCCGCACCGGACCGCGCGCGCGCG 18196
QY 1553 sLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlySe 1573
Db 18197 CGCCCGCGC-----ACGTGCGCGAGGA--GATCTGTGCGGCTTTCGCCG 18243
QY 1573 rLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAl 1593
Db 18244 AGTGTCTGGGACCGGCGCGCATCGACAACTTCTTCGACCTTCGGCGGCACT 18303
QY 1593 aLysSerProHisSerThrValProGluHisHisProHisProIleSerPro-----Ty 1611
Db 18304 CGTGTCTGCCAC-----CGGCTCATCTCCGCGATCCGACCGCTTCGGCGCG 18354
QY 1611 rGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrlsArgSer-----1626
Db 18355 AGATCATCTCGCGGACCTGTTTCGAGGCGCCACCGCTCGCGGACTCGGGAGCGCTCG 18414
QY 1627 -----HisIleProLeu--AlaPheAspProThrSerIleProArgGlyIleProL 1643
Db 18415 ACAGCGTTCGGCGCGCGCGCTCATCCCGCGAGCGCGCGCGGCTCGCGC 18474
QY 1643 euAspAlaAla-----AlaAlaTyrlsLeuProArgHisLeuAlaProAsnProThr 1661
Db 18475 TGTCCTACCGCGCGCGCTGTGTCTTCGCGCGCTCATCCGCGCGGACCGCGCT 18534
QY 1661 yrProHisLeuTyrlsProProTyrlsLeuIleArgGlyTyrlsProAspThrAlaAlaLeuGluA 1681
Db 18535 AC---AACATGCGCGTTCGAGCTCGGCTTCGCGGACCTTCGAGACCGCGCGCGCTG---18587
QY 1681 snArgGlnThrIleIleAsnAspTyrlsThrSerGlnGlnMetHisHisAsnThrAlaAr 1701
Db 18588 -----ACCGCGCGCTTCGCGGACCTCGTGCGC-----CGCCAGAAACCTGCG 18630
QY 1701 hrAlaMetAlaGlnArgAlaAsp-----1708
Db 18631 GCACCGCTTTCGCGAAGCGGACGCGGAGCGCGCGCGCGCGCGCTCGCTCGACCGCGCG 18690
QY 1709 --MetLeuArgGlyLeuSerProArgGluSerSer-----LeuAlaLeuAsnTyrlA 1725
Db 18691 CGCGCGCGCGCGGCTGCAACCGCGGAGCACCGAGGACGACGACGCGCGGAACTGG 18750
QY 1725 laAlaGlyProArgGlyIleIleAspLeu--SerGlnValPro-----1738
Db 18751 CGCGCGCGCGCGCTTCGAGCTTCACACCGAACTGCGCTGCGCGCGCGCGCGCTGT 18810
QY 1739 -----HisLeu--ProValLeu 1743
Db 18811 TCGCGCTCGCGGAGGACGAGACGCTGCTGCTGCTGCTGCGACCATCGCGCGCGAGC 18870
QY 1744 ValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrlsLeuProThr 1763
Db 18871 GCTGTGCTCGCGCGCGCTGCGCGCGGACCTGGTACCGGCTACCGCGCGCGCGCGGAGC 18930
QY 1764 AlaProGlnProPheSerSerArg---HisSerSerSerProLeuSerProGlyGly---1781

Qy	2424	-----GlyAspArgProProSerValSerSerValHisSerGluGlyAspCysAsnAr	2441
Db	20784	CCCGCGCGATCCCGCGCGCGTCC-----	20808
Qy	2441	gA-gThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrPr	2461
Db	20809	-----CCATCGCGCGCGCTGGACACACCGCGCCTACGTCCTCGACGACGCGCTG	20861
Qy	2461	oPheProTyrAsnProLeuMetArgLeuGlnAlaGlyValMetAlaSerPro-----	2479
Db	20862	CGGCCGCTACCGCGCG-----	20891
Qy	2480	-----ProProGlyLeuProAlaGlySerGlyProLeuAla-----GlyProHi	2495
Db	20892	CTGTACATCCCGCGCGCGCTCGCGCGCTACTGGACACGACGCGCGCGCACCGCG	20951
Qy	2495	sHisAla	2497
Db	20952	GAAGCT	20958
RESULT 99			
SPSNBCGEN			
LOCUS			
DEFINITION S.pristinaespiralis snbc and snbDE genes.			
ACCESSION X98690			
VERSION X98690.1 GI:1906377			
KEYWORDS Pristinamycin I synthase 2; Pristinamycin I synthase 3 and 4; snbC gene; snbDE gene.			
SOURCE Streptomyces pristinaespiralis			
ORGANISM Streptomyces pristinaespiralis			
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.			
REFERENCE 1			
AUTHORS de Crecy-Lagard, V., Blanc, V., Gil, P., Naudin, L., Lorenzon, S., Famechon, A., Bamas-Jacques, N., Crouzet, J. and Thibaut, D.			
TITLE Pristinamycin I biosynthesis in Streptomyces pristinaespiralis: molecular characterization of the first two structural peptide synthetase genes			
J. Bacteriol. 179 (3), 705-713 (1997)			
MEDLINE 97158664			
PUBMED 9006024			
REFERENCE 2			
AUTHORS de Crecy-Lagard, V.			
TITLE Direct Submission			
JOURNAL Submitted (21-JUN-1996) V. de Crecy-Lagard, Institut Pasteur, 25 Rue de Dr. Roux, Paris Cedex 75724, FRANCE			
REMARK Revised by [3]			
REFERENCE 3 (bases 1 to 22449)			
AUTHORS de Crecy-Lagard, V.			
TITLE Direct Submission			
JOURNAL Submitted (24-MAR-1997) V. de Crecy-Lagard, Institut Pasteur, 25 Rue de Dr. Roux, Paris Cedex 75724, FRANCE			
COMMENT On Mar 25, 1997 this sequence version replaced gi:1483602.			
FEATURES			
source			
1. 22449			
/organism="Streptomyces pristinaespiralis"			
/mol_type="genomic DNA"			
/strain="Sp92"			
/db_xref="taxon:38300"			
/clone="piBV3 cosmid"			
115..7890			
/gene="snbC"			
115..7890			
/gene="snbC"			
/codon_start=1			
/transl_table=11			
/product="Pristinamycin I synthase 2"			
/protein_id="CAA67248.1"			
/db_xref="GI:1483603"			
/db_xref="GOA:Q54959"			
/db_xref="SPTREMBL:Q54959"			
/translation="MLVPPDDARLPLTAAQLGIWFAQARDPHSPVFNTGEVELTGGLD			

ADRFEAALRATVARTDALRLRHTDAEGTQTQSVADDVPFTLHRVDVSAEDDPAADAA
AMWDVIAERPLDVTGCRPLFTQALFTLPGRRHRWFOHVHLLLDGTFGRFLVRAVAA
YRAGAEAGDAEAGDEPGIRDLIAAEADMRASDAVAQDRAHMLATGTPVTLAGTTA
PGAGALRTASAGPEVLAGLDAAVARAASTGRQVAFPAVAAYTQMLTERTVILGMP
VARTTSAARAPGTVANPVLHVEPRPAELAVSRDDELPLLRHQRYEYSAR
RELGAVERGAYGVVNMVPGYRALDGGPAPASVHPVSTGVEDLKEFVVEEDPGSGAR
IDLEANAAYVGESETDAHLARFVFLGAEAGDPDPLRIHAAAPAAAAAATTGG
PALHTLFAARAAATPHAVAVSHGDHLYAQLDARANLHRLVEOGAGPERLVALTL
PRSLDLVALLAVLSGAAYLPLDDPDYADRLAMWLEDAAPVLVLDPOAMTEDLACY
DIAPRTAVDGAHPAVIYTSGSTVRPGKVVIEHNSNVRLFTSTDHFWGFGDDVDTLIF
HSYADFDSIWEIAGALLHGRVLVVPVHVSPPGDFDLILAREKVTFLNQITPAFHL
DAADRARTAAPELALRYVVGGEALDVARLADMYARRGTAAALNMGYTETVTH
APLPGHVAVGPTSLIGGPIPLDTPRVLDAALRPVPGFTGELYVAGAGTGLVLRNP
ALTAQRFADPYGACGTMYRTGDLVRHLDGGTYVLGRGDDQVKLRGFRILIEGTEN
VLATHPGVAQAANVREDHGRDLRLAAVAPTEGPEPDVAGHLLAGLPAHSLPAS
FTLLAALPUTANGKLDRAALPAQAPAPETADEPAPADPGEAALRRIIADTLGTRPV
GADSPFTLGGHSLLAGRLIAAVDTLGAETI IQDLFEAPTAGLAARVTAATGRPP
LVPRDHGDGPRTSYAQRHMVLRVEGAAPTYNMFLALRLSTLNRALTALRAALYDV
VRHESLRITYPRGEGVTPLLPPEQARPWIESRVDDQLPAALTAARAEFNLDCQ
LPTRAHFLTLCGSDSVLLLMHHVAGDGSRLPLGLDLSTAYAAARLAAAPDPPLPV
SYADPAVQORELLGDHDPDLAARQLAHWESTLAGLPDLQELPADPFRPVAHRGG
SVPLAVPQOHLARLDTFAHAGATCHMWLQALALLRLAGTDLPIGCTASRPDS
ALDDLGVFFNVPLRADTSGDPTFTELLARVATAALAAAHQDVDFRVERLQPER
SAARHPLFVLSYQDFEARHLKGLDVRREDVDFGAKFDLTFLNVEHRGDEEFAG
IDGLEIDYDALYERATAEALTGRLRLLDVAVTDPQAPVGGSHDLLBEAASHARLAFND
TARPVFRAGLAELFTAQARRTPDAVAVGDDGTETLYAELDARSALARRLAAAGAR
ESNAVILQORSVHLVLLAVLKAGGFYVPLNTRYPADRMRLITADVKARVLLTDRAT
DTAHECRSWAGDADVLVDELFTGEGPLVPFAHPQLAYVMYTSSTGNKGVATH
GVVLAADTCWRTGQKRVLLHSPYSFDTQYELMVPLTSGTVVVAAPGDUATRAL
ERAIVDGGVTGMLTSGNLLAESPHCFGTGVEYVMTGGVSVVSPAARVLAASPST
MVADGPTTETTTTATHPMRAPWQOETTPVIGTPLDNTTCHVLDRLRPVPGVTGE
LYIGGSLGARYLNPPELTAERFVADPYAGGRMYRTGDLVRLRRDGVTLFGRAHQ
KVVRGPRIELGIESVLGAHPGIAQOSAVLREDRPGDKRLVAVVPAAGCATATVLR
RHVAALPYMWPAAFVPLDPLTGNKLDRAALPVPDFGGAADSGAPRTPRELL
AVLFAEVLGKRVSDSPFDDGSDSIMVQLVGRARRAGLLNAAQIIEHKTVAOILA
PVAEATLDTTREPAAVATGDVPLTPIMHLERRGGPYDAFSQTMVAVTGLDRER
LVAALQTLIDHDLARAVSRDDADWTMRIEAPGTVDAADVHVHVDAGLDQAALAE
TAAEAAARERLAPFAGGQAVWLADAGHPGRLLTLHLVWDGVSWRILLPLDKA
AMEATAAGSAPALPVVTSFRWAHHLLEQQAGCPEAAELALWTCMLAAGSPVTAQ
PLDGLDTAATAGTLTLLDTTTTAAALLTAPAAENGVRVDDILLTAPALAVTDWRRR
GLGSDAVLTVDEGHGREPGTSGAELSRVGMFTSLYPVRLALGTLDAADALAGPAA
GTAVKRIEALRQIPGGIGHGLRHLRDPRTGSLAALAPYGTPLGFNYLGRFPAG
ADGLMSAAPDAEAI GDGTDPRPLAHAVEVNAITLDEPAGPRLKATWTAGRLDDEPA
VRDLGETWFRALNVTHAERPEAGGITPSDLSLVQLSQDQIDLEAAWKVSP"

7887..22433
/gene="snbDE"
7887..22433
/gene="snbDE"
/codon_start=1
/transl_table=11
/product="Pristinamycin I synthase 3 and 4"
/protein_id="CAA67249.1"
/db_xref="GI:1906378"
/db_xref="GOA:O07944"
/db_xref="SPTREMBL:O07944"
/translation="MTRSLADVLPLSLQEGLLFQVITYGAETHGPDVTVVMVFELR
GPLDERGLAAALILRHPLRAGFMQOGVERPVQFVNEVPLPHHTDFTALGESD
REBAVEAYTADAEERFPDGPAPLIRFALLGLAADHKLVLTHLLLDGMSMLVLR
ELFTLYGQRDDAGMPFPTPAYLAHLAGRDDDAARAARALTADLERPSLVAGAGA
GRGAADGSALPGQIWEIDAATATTALARSCTITLQHPGAKRLGTAAARPOLGRDDV
VFGATVAHRPEIPIGIESTIGINTLTVRVVRPAETLGDLLGRVQREQAALIEHRH
LSLTDIRSTGTGELFDTVVFPENYPLDPAVLRARAGRLAGFVSDATHYPLSLILA
IPGESLRPLTDHRGDVDEAGALLERLDTLLTDIAEHGADLPVGRDILLGSAERHQ
VLEENDTGLPAEDATLALFEAQARTPDTALLVGGSLTYAEALNARLARIHLV
TLGAGEQIVAVKLESLDLYVALLAVLKTGAAYLPDVTAYFAERI AFMDDARPAV
LTGEDTQDLSGVDLTDLDATPLPAHPAYVIYTSGSTGTGTPKAVMPPGALVNL
LAHRRRETFEGAGVPAQPTTIGFDVAAQETILATLWHKLTAVPSQEVRSRSLAA
LDEQHVSELYAPNLVIEALAAEAAAGRTLPALRHIAQAGEALTTRTREFAAAPG
ROLHNHYGPAETHVMTGTALPEDPASEHAPLGRVPSGARVVVLSALRPVAPGVYTG
ELYLAGAGVSRGYLNPVLTAEFVADPYAPSGARMYRTGDLGRWNTGEGLEFAGRA
DHQVKIRGPRIRIEPEIEAALTDLPVAAAANVVRERPGDKRLVAVAPAGSGLDAA
VSRRLARTLUPDFWPAALVALLDALPLTNGKLDRAALPAQSTTAVQRTIPRSPREIL
TTLFAAVLKPVRGIDDDNFFDLGGHSLALRLISRTVILGAEITLURDLFEAPTAVGL

GERLDQASGARPALTPABRPERLPLSHAQRRLWFLGRLDGPNSTYNIPLALRLSGTLD
TAALRALADLVARHETLRTYPSHDEQYQHIVAPADATPALTVVTHEERLAEOLA
EATAPYDLTTELPLRLTFRIGDEHEVLLHLLHAGDWSLAPITRDLAAAYAAAR
DGHAGPAPLPVOYADYTLWQGLIGDADDPSTLOAAQLAHWRQALEPRAHLEPCD
HPPAVATHRGATVPPEHIDAGLEKLALAKACDSSLFWULQAAPALLTRHGAGDPI
PVGSPVAGRTDADLDDLVGFVNTVLURTUTSGDPTFRELVARVQFDLAAYTHQDMP
FEKLEBVENPERSLARNPFLQVVALQSPMAADLAUPLDVAAPVVRVGFPAKFDGLA
VIEBHATDGTFRAGIWDWSTELFERGTVEALGSLRLLDADVPDPTIGSVOLL
DATERHLLVGLANTDQPLPEERTLTALFEQAAATPDATALVWGRDSTLYAELDARA
NRLARHLHAGCEQIVALQOLPSRLDILTAVLAWKSAAIYLPVDPDYPPIAHMR
ADAPALVDAI PDDTTLAAYADSRLLTDADESAPLLPAHPAYVITYSGTGAPKGVV
AHSLSAATVPAQAAGFGTHSRVINFASISFDAVWELISALTGAGLVADADLL
PGPSLARHRRHITLALPSALPALPDGALPGTDLIVAGDATAPDQAARFAPGR
MVNAYGLTETTCATSEPAFGDAPP IGRFVAHARVYVILDERLVRPVPCTGEMATVS
GAGVAGYLRPLTAQFQFVDPVALLFGSTGTMYRTGDLARLDADGRHLRFGTGRADY
QVKIRGPRIPBGEIETVLTAPPAVAGAVIAREDTGDKOLVAVLTRDTTHAAPDQD
LGTCQVDRWQOTDSLYEAEPGAFGEDFSGNSSYTAEBIPIEBMRWRAATVDRVL
ALAPRRVLETGCGTGLSVQAPHTSEIRGTDLSRVSVAARLAHLRAGPLADKVTVD
ARAHAETDLPKGRFDTIVLNSVAQYFPDARYLAGILHRAAELLAPGDTFLGDIRNL
RTLHAETFAVELRRAGFADPAARVRAVEQSALTEKELLLDPPDFTDLVDRLPHLTRA
DITLRTAAHNENSRHRYDVLTRHGGDSPAPAEVTLRWGTDLDTAGELARGLVSGRPV
RVGTINARLITGTRAAQALERSAADACARLEAPAGALDPHDVVALGAVCTWGD
DDTFDAYAGGTAYRPTGCGPLANDPARAQODAAALVGBRLDLVQRLPAHWAPAY
VLDRPLSGANGKLDRLAPDGRDGAQAPTRREBIILSTLFAEVLGPAVGDIR
SFDLGHSLLLATLSRISIRSVLGAELAVRDLFOAPTAAALAEVRVGAQAPPLTAR
TRAEVPLSPAQYRLWFLHRMEGFGPTNYI PMSLRITGLTDLTAAALAAADVTARHEA
LRTVYPSGSGVQPHVLHDVHPHFETADTTAQTLEADLTTRAARRGDLARELPRLATL
FRLAEDHEVLLHMHITAGDWSMPLPLARDLGDVAYARLAGRAPDLPPLPVQYADVAL
QDRLDGEDDPTSRHARQLAYWAGALDDLPEBLQPLCDRPRPAVATHRGDTVPFVAVD
PDLHQRNLATGSRASLFWLQAFAALLTRHGAGDITPLGSPITAGRTDADLLELVG
FFVNTLVRTDTSQDPTFRELQVRRESLDAAYAHQDVPPEKLVKXLPERSLSRPL
FQVMLAFQNSGEARLALGLDARALPVGVAKFPLDLHLSNVELRTDGRPGGLOAALE
YATDLDFRTAQLVRLRLAAVADDPGLTVGRVELLDPAERHRLPDLODNTPGDID
GAARTPVTELLRQOARTPDVALVHDDGRLTYAEULHARANRLARHLITLGAPEQIV
ALRMPRLDLYVALLAVLTKAAYLPVDSIYPAERIAPMEDARPVTVLDRLPDDLGA
VRDUTDADATGFPSPPEHPAYVIHTSGSTGTPKAVMPHAGLANLTHWARRPGGT
GVRTAQTATGFPVQVEILSPVMGKTLAVPSEVRHSAELLAGLWETQINELFAP
NLVIEALAAEABAGRTLPDLTLOQGEALAPTRVAPTAAPVGRHNLVYGPART
HAVTHTLAADPAHWPDSAPIGRPVHDVRLDSALRPVPGVTGELYLAGAVARG
YLHRPRTAERFVADPVAPRPGARMTVTGDLGRWNTGGELEFAGRADHOKVIRGFVE
PGEVEAALTPRITQMAVLHAGDRLVAYVYVTTADRLDTGVRHLEALRLPDMVPTAY
VRUEALTPVNGKLDRAALPAPAGRAPRTVREBILGCLFAEVLGTERAGIDONF
FDLGHSLALTRISIRLTALGABITLRLDLEAPTAVAGLERLDQASGARPALTPABR
PERLPLSQAQRLLWFLRGLGDPGYMPVPLSGTLDTAALTAALADVARHETLR
TVPEGDSGPHOLVDPGAARPALPRDPTDQDLATLAAAHACHFDLTTELPLRAT
LFLRAEDHEVLLHLLHIIAGDWSLAPLARDIATAAARDGHAPDAPLPVQYADYT
LWQHALLGDENDPSTISROSAHWKALAGLPEEIQLPADRPAPPAHREGAVFTL
DADHRELLHARTGASLFWLQAAVAVLTRHGAGDITPLGSPITAGRTDADLDDLV
GFFVNTLVRTDTSQDPTFRELQVRRESLDAAYAHQDVPPEKLVKXLPERSLSRHP
LFQVLLALQNTPEAVLELPGLTARPEVTLGAAKFDLTFLNLAERHGPDPGPHGIDAVL
EYSTDLDFRTTALTRLTFLRAVAADPGRTVADVPILDADHARVLTGWNDDOHP
RGRTADACLPRIAEQAAKTTHAVATEAGGTLITVSELDARANRLARHLIRRGVTA
ETRVAVLAERSAQVLVTLTALIKAGVYVPLHTGYVDRMRHVLDAEALLTDTTH
AATAARDTALTVDTEITAGTQDTPADVPDQVLAIFYMFTSGSTGTPKGIITGTHR
DATALAADRCMDLDSRLVMSHYPADITSTFELFAPLAGRIVAPRGDIDAAVQV
RTLAAGHVTSLLTAGLLGVIADAEPEVFTSGVDWMTGVDVSPVAVRVLACPGTV
VKTLGYTETTLTLPFTDPRRIIPVAPIGRPLDNTRAYVILDERLVRPVPCTGEL
YIAGLAGRCVQDSQARTAEFTADPAHLFGDTGGMVYRTGLARDADGVLFQGR
ADQVKIRGRIRIEGEIETALAHPDVTRAARVAPRAGDKVLVAYLVATPDAGDTT
AEQURAGLETQLEPDMVPAFVALPALPVTVPNGKLDRLDLPEDPDWGGGPPRGPRE
EILCLGFPAEDPAPYVTDNDTFFELGSHSMATLVRVKTIVLAGDITGVRLFEATV
AALAAARDGADTGHDPIDVVLPLRTGSRAPLFCVHPAGFGVTLGWSFGSLVAHAAVROPL
YALQAARGLARPELPDDIDAMARDYAEQIRKTVPEGPYEITLGMWSFGSLVAHAAVROPL
AEGASGLSLAVLDGYPADYDTEHEVGEQVLAILLNAGVDRQAQFQDAPLQRAAVL
EKURDSGLAGNLDHDDAVGRVTVFLNTRLIQNFRPRTGDMVTFPAAAGHHDPAL
TPGNWRPVTYVTRIEEHLDDTHDGLARPEALGAVARTLAQRDTPQAGH"

ORIGIN

Alignment Scores:

Pred. No.:	5,246-05	Length:	22449
Score:	654.00	Matches:	683
Percent Similarity:	31.26%	Conservative:	290
Best Local Similarity:	21.94%	Mismatches:	1193

Query Match:	4,95%	Indels:	959
DB:	1	Gaps:	136
US-09-522-753-5 (1-2517) x SPSNBCEGEN (1-22449)			
Qy 8 ValAlaGlnThrTrpArgAla-----ThrGluProArgTyrProPro 21			
Db 12636 GTCTCGCGGTGTGAAGTCGCGCGCGCTACCTGCTCGCGCGGACTACCGCC 12695			
Qy 22 HisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuGlu 41			
Db 12696 GACGGCATCCCCACATGCGCGCGGACGCGCGCCCTCTGCTGTCGAGCGGATCCCC 12755			
Qy 42 TyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleLeuGln--- 60			
Db 12756 GA-CGACACCCACGCT-----GGCGGCATACGCGCGACG 12787			
Qy 61 ---ProGlnArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSer 79			
Db 12788 CCGCTCTACCGACGCGCGACGCGCGCGCTGCTGCTCGCGCCA-----CCCGGCTA 12841			
Qy 80 GlnGluLeuHisLeuArg-----ProGlu 87			
Db 12842 CGTCATCTACCTCCGCTCCACCGGGGGCGCCCAAGGGCGTCTGCTCGCCACCGCTTC 12901			
Qy 88 SerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArg 107			
Db 12902 ACTCGCGCGGACCGTCCCCCGCCCGCGCGCTTCGGACTGGGCACCCACGCGCGT 12961			
Qy 108 ProArgLeuGluLeu----- 112			
Db 12962 CTTGAACCTTCGCTCCATCAGCTTTCGACGCGCGCGCTGCGAGCTGACCTCCGCGCTGT 13021			
Qy 113 -----LeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGly 127			
Db 13022 CACCGCGCGGAGCTGCTCTGCGCGACGCGCGAGCTGCTGCTCGCGCGCGCTCGCTGGC 13081			
Qy 128 GlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluPro 147			
Db 13082 CCGCTCGTCCACGACGCGGCACATCACCCTGATCGC-----CCTGCGCGCGCTCGCGCT 13135			
Qy 148 ValSerPro----- 150			
Db 13136 GCCCGCGCTGCGCGCGCGCGCTGCGCGCGGACCGAGCTCATCTGCTCGCGCGCGACG 13195			
Qy 151 -----ProSerProProHis----- 155			
Db 13196 CACCGCGCGGACGACGCGCGCGCTTTCGACCGCGCGCGCGATGCTCAACGCTACGG 13255			
Qy 156 ThrAspProGluLeuGluLeu----- 162			
Db 13256 ACTGACGAGACACCGTGTGCGCCACCATGAGGAGCGCGCCACCGCGCGCGCGCGCC 13315			
Qy 163 -----ValProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp 180			
Db 13316 GCCCATCGCGCGCGCGCGCGCGCGCTGCGCGCGCGCTCTACGCTCTGAGACGAGCGTCCG 13375			
Qy 181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnLeu 200			
Db 13376 CGTACCGCGCGCGCGTCAACGCGGAGATGATGCTCTCGCGCGCGCGCTCGCGCGCGCTA 13435			
Qy 201 GluGluGluAlaAlaLysProGluProGluPro-----LysProVal 214			
Db 13436 -----CCTGACCGCGCGCGCGCTGAGCGCGAGCGCTCTGTCGCGCGCGCGCTA 13483			
Qy 215 SerProProProIleGluSerLysHisArgSerLeuVal---GlnIleIleTyrAspGlu 233			
Db 13484 CGCCCTGCTGTTCCGCGGAGACCGCGCATGATACCGCACCGCGCGACCTGCGCACGCT 13543			
Qy 234 AsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyProGlnValGlu 253			
Db 13544 GGACGCGCGCGCGCTGCTGCTTCGCGCGCGCGCGCGCGACGACGAGGTCAAGATTCGCGG 13603			

QY 254 LeuPro-----LeuTyrAsnGlnProSerAspThrArgGln 265
|||
D 13604 CTTCCGATCGAACC CGCGGAGATCGAGACCGTCTGACCGCCACCCCGCGCTGCCGC 13663
QY 266 TyrHisGluAsnIleLysIleAsnGlnAlaMetArg-LysLysLeuIleLeuTyrPheLys 285
|||
D 13664 CGCGCGGTGATCGCGCGGAGACACCCCGGGCGACAAACAGCTCGTCTACCTGAC 13723
QY 285 sArg-----ArgAsnHisAlaArg-----LysGln 293
|||
D 13724 GCAGACACCCACCCACCGCGCGGACCGGACCGGACCGGACCGGACCGGACCGG 13783
QY 293 nTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysVa 313
|||
D 13784 CTGGCAG-----CAGACCTACGACTCCCTCTACGAGCGGAAACCCCGCGCGCG 13831
QY 313 l-----GluArgIleGluAsn---AsnProArgArgAlaLysGluSerLysValAr 330
|||
D 13832 CTTCCGCGAGGACTTCTCCGCGTGGAGACAGCAGTACACCGCGGAGAGATCCCTGGA 13891
QY 330 gGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGluArgMe 350
|||
D 13892 GGAGATCGCGAGTGGCGCGCGCCACCGTCCGACCGCGTCTCTCGCCCTGCGCCCGCG 13951
QY 350 tGlnSerArgValGly----- 355
|||
D 13952 GGTCTGGAGATCGGCTGGCGCACCGGCGTGTCTGTCACAGTTCGCCCGCCACACCGA 14011
QY 356 ---GlnArgGlySerGlyLeuSerMetSerAla----- 365
|||
D 14012 GGAGTACCGGGGACCGACCTCTCCGACGCGTCTGCGCGGCGGCGGCGGCGGCGG 14071
QY 366 -----AlaArgSerGluHisGluValse 373
|||
D 14072 CGGCGCGCGCGACCTGGCGGACAAAGTACCGTGGACCGCGCGCGGCGGCGGCGG 14131
QY 373 rGluIleLeuAspGly-----LeuSerGluGlnGluAsnLeuGluLys 387
|||
D 14132 CGACTGCGCCMAGGCGCTTCGACGATCGTCTCACTCGTCCGCGCGGCGGCGGCGG 14191
QY 387 sGlnMetArgGlnLeuAla-----ValIleProPro----- 397
|||
D 14192 CGACGCGCGTACCTCGCGGATCTCTCCACCGGCGCGCGGAACTGTGCGCGCGGCGG 14251
QY 398 -MetLeuTyrAspAlaAspGlnArgIleLysPheIleAsn----- 411
|||
D 14252 CACGATCTCTCGCGCGACATCCGCAACCTGCGCACCCCTGCGACCGCTTCGCGCGCG 14311
QY 412 -----MetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnVa 428
|||
D 14312 CGAACTGCGCGCGCGCGCGCTTCGCGGACCGCGCGCGCGGCGGTA-----CGCGCGC 14362
QY 428 lMetAsnMetTrpSerGluGlnGluLysGluThr-----PheArgGln 442
|||
D 14363 CGTCAACAGTCACTGGCGACGAGAGAACTCTGTCTGACCGCGGCGGCGGCGGCGG 14422
QY 442 uLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysTh 462
|||
D 14423 CCGTGTGACCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 14473
QY 462 rValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLe 482
|||
D 14474 GGCGCGCGCG-----CACAAACGAGATGAGCGCGCGCGG 14506
QY 482 uValArg----- 485
|||
D 14507 -CTACGAGCTACCTCCACCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 14565
QY 486 -----SerTyrArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGln 500
|||
D 14566 GCTGGGCGACCGACCTCGACACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 14625
QY 500 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520

D 14626 CCGTGGCGCTCACCGCATCCCAACGCGCGCTCACCGCGGACACGCGCGCGCACAA 14685
QY 520 sAsp-GluLysGluLysGluLysGlu-----AlaGluLysGluGluGluLysProGluV 538
D 14686 CCGTCAACCGGATCGCGCGCGCGCGCGCTCGCGCGCGCTCGAAGCCCCCGCGCGCG 14745
QY 538 alGluAsnAspLysGluAspLeuLysGluLysThrAspAspThrSerGlyGluAspA 558
D 14746 GCGCACTCGACCCCGACGACGCTGTACGCGCTGGCGCGCTGTGCACCTGGGCGCACG 14805
QY 558 snAspGluLysGluAlaValAlaSerLysGlyArg-LysThrAlaAsnSer---GlnGly 576
D 14806 ACACACCTTCGACGCTACGCGCGCGCGCGCGCTACCGCGCGCGCGCGCGCGCGCG 14865
QY 577 ArgArgLysGlyArgIleThrArg-----SerMetAlaAsn 588
D 14866 AGGAGCCGCTCGCCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14925
QY 589 GluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAla----- 605
D 14926 TCGCGGACTGTCTCGCCGAGCGG-CTGCCCGCCACATGCGCGCGCGCGCTACGCTCG 14984
QY 606 -----SerMetGluLeuAsnGluSerSerArgTyrTrpThrGluGluMetGluThra 623
D 14985 CTCGACCGCGCTCGCGCGCAACCGCAACTCGACCGCGCGCGCTGCGCGCGCGCG 15044
QY 624 LysLysGlyLeuLeuGluHisGlyArgAsn-----Trp 634
D 15045 GACCGCGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15104
QY 635 SerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheTyrPhe 654
D 15105 ACCCTGTTCCGCGAGTGTGCGCGCTGCGCGCGCTCGCATCGACCGCGCTTCGCA- 15163
QY 655 AsnTyrLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGlu 674
D 15164 CTTGGCGCGCACTCGTTCGCGACCGCGCTGTCTCAGCGCGCTCGCGCGCGCTCTGG 15223
QY 675 LysGlu-----ArgAsnAlaArgArgLys 682
D 15224 CGCGGACTCGCGCTCGCGACCTGTTCCAGCGCGCGCGCGCGCGCGCGCGCGCGAG 15283
QY 683 LysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAsp 702
D 15284 CGTCGACGCGCGCGCGCGCGCG-----CCGCGCG 15310
QY 703 GluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaGlu 722
D 15310 ----- 15310
QY 723 AlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrVal 742
D 15311 GCGCTCAACG-CCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT- 15351
QY 743 AsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThr 762
D 15352 -----CCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15384
QY 763 GlyGlnAsnGlyProLysProAlaThrLeuGlyAlaAspGlyProProGlyPro 782
D 15385 GCATGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15444
QY 783 ProThrProProArgArgThrSerArgAlaProIleGluPro----- 796
D 15445 TGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15504
QY 797 -----ThrProAlaSerGluAla-----ThrGlyAla 805
D 15505 GCACCGCTTACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15564
QY 806 ProThr-----ProProAlaProProSerProSerAlaPro-----Pro 819
|||

Qy	1371	LeuArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProSer	1390
Db	17634	-----	17654
Qy	1391	Arg-----AspLeuThrGluAlaTyLysThrGlnAlaLeuGlyProLeuLysLeuLys	1408
Db	17655	CGCCCGTGCACACGACCGCGTGTACGTCCTGGACAGC-----GCACTGCGC	17702
Qy	1409	ProAlaHisGluGlyValAlaThrValLysGluAlaGlyArgSerIle-----	1425
Db	17703	CCCGTCCCGCGCGCTACCGCGAGCTGTACCTGCGCGCGCGCGCGCTCGCCGCGC	17762
Qy	1426	-----HisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaPro	1443
Db	17763	TACCTGCACGCG-----CGCGCACACCGCGAGCTTCGTGCGCGACCCCTACGCCCA	17819
Qy	1444	ArgPro-----LeuLysGluGlySerIle-ThrGlnGlyThrProLeuLys---	1458
Db	17820	CGCCCGCGCGCGCATGTACCGCACCGCGACCTCGCGCGTGGAAACACCGCGCGGAG	17879
Qy	1459	----TyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeu1	1477
Db	17880	CTGGAGTTCGCGCGCGCGCGCGCACACAGGTCAAGATCC-----GCGCTTCGCG	17930
Qy	1477	eGlySerProGlyArgThrPheProValHisProLeuAspValMetAlaAspAlaAr	1497
Db	17931	GTGGAGCGCGCGAGGTTCGAGCGCGCGCTCACCGCGCCACCGCGCATCACCCAGGCGG	17990
Qy	1497	gAlaLeuGluArgAlaCysTyArgGluSerLeuLysSerArgProGlyThrAlaSerSe	1517
Db	17991	GTCC-----TCGCGACGCGCGCGCGCTCGTC	18017
Qy	1517	rSerGlyGlySerIleAlaArgGlyAla-----ProVa	1528
Db	18018	GCCTACGTCTACACCGCGGACCGCGACCTGACCGGGTACCGAGCACTGCGCGCGC	18077
Qy	1528	lIleValProGluLeuGlyLysPro-----ArgGlnSerProLeuTh	1542
Db	18078	CGCTGCGCGACTTCATGTGTCCCAACCGCTACGTCTGCGCTCGAGGCGGTG-CGCGTCA	18136
Qy	1542	rTyArgLysHis-----GlyAlaProPheAlaGlyHi	1553
Db	18137	CGTGAACGGCAAACTCGACCGGCGCGCTGCGCGCACCGGACCGCGCGCGCGCGG	18196
Qy	1553	sLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlySe	1573
Db	18197	CGCCCGCGCG-----ACGTGCGCGAGGA-GATCTGTGCGCGCTGTTCGCGC	18243
Qy	1573	rLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAl	1593
Db	18244	AGGTGTGGCACCGAACGCGCGCGCATCGACGACAACTTCTTCGACCTTCGCGGCACT	18303
Qy	1593	AlaSerProHisSerThrValProGluHisHisProHisProIleSerPro-----Ty	1611
Db	18304	CGCTGTCTCGGCAC-----CGCGCTCATCTCCCGCATCCGACCGCGCTCGGCGCG	18354
Qy	1611	rGluHisLeuLeuArgGlyValSerGlyValAspLeuTyArgSer-----	1626
Db	18355	AGATCACCTCGCGGACCTGTTGAGGCGCGCACCGTCCGCGGACCTGCGGAGCGGCTCG	18414
Qy	1627	-----HisIleProLeu-AlaPheAspProThrSerIleProArgGlyIleProL	1643
Db	18415	ACCAGGCGTCCGCGCGCGCGCGCTCACCGCGCGAGCGCGCGCGCGCTCCCGC	18474
Qy	1643	eAlaAlaAla-----AlaAlaTyTyTyLeuProArgHisLeuAlaProAsnProThrT	1661
Db	18475	TGTCTTACGCGCGCGCGCTGTGTGTCTTCGCGCGCTCACCGCGCGCGCGCGCT	18534
Qy	1661	yrProHisLeuTyTyProProTyTyLeuIleArgGlyTyTyProAspThrAlaAlaLeuGluA	1681
Db	18535	AC-----AACATGCGCGTGGAGTGGCGCTGTTCGCGCGCGCGCGCGCGCTG---	18587
Qy	1681	snArgGlnThrIleIleAsnAspTyTyIleThrSerGlnGlnMetHisHisAsnThrAlaT	1701

Db 19596 GCGGCCAGTTCGACCTCCTCAACCTGGCGGAACGGCAGCCGCCCGGAGAACCC 19655
 Qy 1972 ----AlaSerSerProSerLysGlySerGluProArgProLeuValProProValSerG1 1990
 Db 19656 CACGGCATCGACCGCGTCTCGAATACAGCAGCAGCGCTTCGACCGCACCACCGCGGAA 19715
 Qy 1990 yHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAs 2010
 Db 19716 GCACCTCACCGACCGGCTCATACCT-----TCCTGGCGCGC-----TCGCGGCC 19760
 Qy 2010 pProProAlaProProAlaSerAla-SerAspProHisArgGluLysThrGlnSerLysP 2030
 Db 19761 GACCCCGCGCGACCGTTCGCGGAGTCCGATCTCGACCGCCAGCAGCAGCGCGGTC 19820
 Qy 2030 roPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyHisGlySerSerTySerP 2050
 Db 19821 CTCACCGGCTGGAACGA-----CACCGACACCCCGCGGC 19856
 Qy 2050 roGluGlyValGlu-----ProValSerProValSerSerProSerLeuThrHisAspL 2068
 Db 19857 CGCGCGCACCGCGCTGCTGCTCCCGCGCATCGCCGAACAGCGCGCGCCGACCCCG 19916
 Qy 2068 yGlyLeuProLysHisLeuGluLeuLysSerHisLeuGluGlyGluLeuArgP 2088
 Db 19917 CACCGGTCGCGTCCACGAAGCGCGCGGACCCCTGATCCTACCTACCGAAGAACTCGACGCC 19976
 Qy 2088 roLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgP 2108
 Db 19977 CGCGCCACCG-----C-----C 19988
 Qy 2108 roLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaPro-GlyValLys 2127
 Db 19989 CTGCGCGCGCACCTGATCCG-----CCGCGCGGTCAACC 20021
 Qy 2128 GlyHisGlnArgValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTy 2147
 Db 20022 GCGGAGACCGCGTCCGCTCGCGGAACGCTCGCGCCAGCTCGT----- 20068
 Qy 2148 ThrArgHisHisProGln-----GlnLeuSerAlaProLeuProLeuTySer 2165
 Db 20069 ---CGTCACACCTCGGATCTCAAGCGCGCGGCGGTGTAGTGCCTCGCACCGG 20125
 Qy 2166 Phe-----ProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAsp 2182
 Db 20126 CTATCCGTCGACCGGATCGGCAGCTCTCGCGCACCGAAGCGCGCTGTGTCTCAC 20185
 Qy 2183 LeuTyLeuPro-----ProProAspHisGlyAlaProAlaArgGlySer 2197
 Db 20186 CGACACCCACCGCGCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20245
 Qy 2198 ProHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGly 2217
 Db 20246 GGACACCA-CGGCGCGGCGACGACACACCGCGC---CGGAGTTCACCGTCCGCGC--- 20298
 Qy 2218 GlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSer 2237
 Db 20299 -----CGACACGAGTCCGCTACATCATGTTCACCTCCGCT----- 20334
 Qy 2238 ArgSerAlaValTyProLeuLeuTyArgAspGlyGluGlnThrGluProSerArgMet 2257
 Db 20335 -----CGACCGGACCCCAAGG--- 20352
 Qy 2258 GlySerLysSerProGlyAsnThrSerGln-----ProProAlaPhePheSerLysLeu 2275
 Db 20353 GCATCGGCATCACCCCGGAGCGCATCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 20412
 Qy 2276 ThrGluSerAsnSerAlaMetValLysSerLysGlnGluLeuAlaSerLysLysLeuAsn 2295
 Db 20413 ACACCGGCTCCCGGCTCATGCATCTCCCGTACGCTTCGATCTCCACCTTCGAAC 20472
 Qy 2296 ThrHisAsnArgAsnGluProGluTyArgAsnIleSerGlnProGlyThrGluIlePheAsn 2315
 Db 20473 TGTTCCCGCGCTGCTGGCGCGCGGCGGATCGTCC-----TCG 20511

Qy 2316 MetProAlaIleThrGlyThrGlyLeuMetThrTyArgSerGlnAlaValGlnGluHis 2335
 Db 20512 CACCCCGCGCGACATCGACGCG-CGCGTCTCTGAGCGCACCTCGCGCGC-----CAC 20564
 Qy 2336 AlaSerThrAsnMetGlyLeuGluAlaIleAlaLeuArgLysAlaLeuMetGlyLysTy 2354
 Db 20565 GCGGTACACCTCGTGTCTGTCGACGCGC-----GGCCTGCTCGCGGTGATCGCC 20612
 Qy 2355 AspGlnTrpGluGlu-SerProLeuSer-----AlaAsnAl 2367
 Db 20613 GACGAGCAGCGAGGTCTTCACGCGGCTCAGGACGTGGACCGCGCGCGCGCTCGTC 20672
 Qy 2367 apheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAs 2387
 Db 20673 TCTCCACCGCGCTGCGCGCGCTCTCGAAGCCTGCCCGCGCACCG----- 20718
 Qy 2387 pGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerG1 2407
 Db 20719 -----TCGTC 20723
 Qy 2407 YArgProSerSerArgLysAlaLysSerPro-----AlaProGlyLeuAlaSer----- 2423
 Db 20724 AAGACCTCTACGCGCGCGCGGACGACACCTCGGCTGCGACCTCGCTCGCGCTTCACCGAC 20783
 Qy 2424 -----GlyAspArgProProSerValSerValHisSerGluGlyAspCysAsnAr 2441
 Db 20784 CCGCGCGCATCCCGCGCGCGTCC----- 20808
 Qy 2441 gArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrPr 2461
 Db 20809 -----CCATCGCGCGCGCGCTCGGACACACCGCGCGCTACGCTCTCGACGACGCGTG 20861
 Qy 2461 oPheProTyAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerPro----- 2479
 Db 20862 CGCGCGGTACCGCGC-----CGTCACCGCGGAG 20891
 Qy 2480 -----ProProProGlyLeuProAlaGlySerGlyProLeuAla---GlyProHi 2495
 Db 20892 CTGTACATCGCGCGCGCGCTCGCGCGCTACTGGGACGACGAGCGCGCGCACCGCG 20951
 Qy 2495 sHisAla 2497
 Db 20952 GAACGCT 20958
 RESULT 100
 SC0939114
 LOCUS
 DEFINITION Streptomyces coelicolor A3(2) complete genome; segment 11/29.
 ACCESSION AL939114 AL096852 AL138058 AL138851 AL138852 AL163641 AL163672
 AL353832 AL353862 AL355774 AL442165 AL499570 AL596030 AL596138
 AL645882
 VERSION AL939114.1 GI:24413773
 KEYWORDS
 SOURCE Streptomyces coelicolor A3(2)
 ORGANISM Streptomyces coelicolor A3(2)
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Streptomycineae; Streptomycetaceae; Streptomycetes.
 REFERENCE 1
 AUTHORS Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L.,
 Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H.,
 Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M.,
 Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornby, T., Howarth, S.,
 Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neill, S.,
 Rabinowitz, E., Rajandream, M.A., Rutherford, K., Rutter, S.,
 Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S.,
 Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrall, B.G.,
 Parkhill, J. and Hopwood, D.A.
 TITLE Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2)
 JOURNAL Nature 417 (6885), 141-147 (2002)
 MEDLINE 2196410
 PUBMED 12000953


```

/db_xref="SPREMBL:Q9L069"
/translation="MSLADAVAHILTPERWEANRLVVKALAEPTHERLLTPEREPPD
GGGYVVRSDGQTAYRFTATRALDHQVDAASVTRHRDGAELPLAALDFELKQ
TLGLSDILPYLEIEISITLSTGTCYKLTQKLSARLARSDDFOAVETGMEHPCFV
ANNGRLFGFHEVLSYAPETASVRLVLAHRSRAAFTAGVGLYESFVRDELGAAT
VDRPHVLRGRLDPADYLLIPHPWQWKNLITVFAEVARGHLVCLGEGDEYLAQ
QSIRTFNAPHPGHYKVTALSYNMGMFRLGSLSAAYMEATPAINDWLARLIEGDPVLK
ETGLSITREAAAGVGRHLEYQATDRSPYRKMLAALRESVPFSIREGETLATMASL
VHODHEGASGALIERISGLTPTTEHLYRAYVPLLSFYADLYVMPHGENVILV
LADQVVRVVKDIAEBIAMDPAVLPPESVRIADVDPDKLLSTFTDVFDCFRPF
LAANLAEGIVTDEAFWRTVAETVREYQESVPELADKFEYDMFAPFALSCLNRLQ
RDNQVMDLADPSGALQLVGTLKNPLAGR"
/gene="SC02786"
/complement(6302..7909)
/note="synonyms: hexA, SCC105.17c"
/complement(6302..7909)
/gene="SC02786"
/note="SCC105.17c, hexA, beta-N-acetylhexosaminidase
precursor, len: 535 aa; highly similar to TR:085361
(EMBL:AF063001) Streptomyces plicatus
B-N-acetylhexosaminidase, Hex, 561 aa; fasta scores: opt:
2991 z-score: 3318.1 E(): 0: 91.2% identity in 469 aa
overlap and to TR:BA08876 (EMBL:AB015350) Streptomyces
thermophilaceus beta-N-acetylglucosaminidase precursor
NagB, 541 aa; fasta scores: opt: 2856 z-score: 3168.6 E():
0: 76.7% identity in 537 aa overlap. Contains pfam match
to entry PF00728 Glyco_hydro_20, Glycosyl hydrolase family
20 and possible N-terminal region signal peptide sequence"

```

gene

CDS

Alignment Scores:

```

Pred. No.: 0.000437 Length: 313800
Score: 654.00 Matches: 637
Percent Similarity: 31.34% Conservative: 264
Best Local Similarity: 22.16% Mismatches: 1051
Query Match: 4.95% Indels: 935
DB: 1 Gaps: 126

```

US-09-522-753-5 (1-2517) x SC0939114 (1-313800)

```

Qy 21 ProHis---SerLeuSerTyProValGlnAlaArgThrHisThrAspValGlyLeu 39
Db 113934 CCGACCTTGGCTTGACACACAGCTCCAGCCGATCGTCGGCTGTCAGAGTACAGCAC 113993
Qy 40 LeuGlu---TyrGlnHisHisSerArgAspTyAlaSerHis---LeuSerProGlySer 57
Db 113994 CTCGGGTCTGTCAGCAGTGCCTGGGATCTGCGCGCATCGCTGCC-CCAGGGAGA 114052
Qy 58 IleGlnProGlnArgArgProSer-LeuLeuSerGluPheGlnProGlyAenG1 77
Db 114053 GTTCCCGCACCGCAGCTGCGAGCAGCTCCTGGAGCTGGAGAGTTCGACACAGCGGTCCA 114112
Qy 77 uArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyLeuLeuProGluLeuGly 97
Db 114113 GGTTCCTCGGTAAACGGGGTCCGGGATGCGGTATGCGGTGATCGCGGTAGGAGT 114172
Qy 97 sSerGluMetGluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLe 117
Db 114173 CGATCAGCGCAGGTCCACACAGCGTCTGCGCTGCCGACACACACCGCGGTGCGGT 114232
Qy 117 uLeuArgProSerProLeu-----LeuAl 125
Db 114233 GCGCCAGCGGATCCGCTCCGGGACGGTTCGATGCCCGCACCGCGCGCGCGC 114292
Qy 125 aThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArg----- 139
Db 114293 TCGCGGTGAGATGCGCGT---GAGCATCTGTGATCGTCTCCCGCGCGCGGTTCG 114349
Qy 140 -----SerLeuThrGlyLysLeuGluProValSerProProSerProPr 154
Db 114350 GCGCGATGTAGCCACCATCTCGCGCGCGCACCGGTGACGACGAGCGCGTCCACCGCCC 114409
Qy 154 oHisThrAspProGluLeuGluValProProArgLeuSerLys-----GluGluLe 172
Db 115284 CCACCTCGGGCAGTAGTACCGCGCGCGCGCGTCCACCGCGGAGAACATCAGCAGGATCG 115284

```

```

Db 114410 GCACCTG---CCGCGCGTCTCGGCTTTCAGGAAGCCCGCTTCTCTCGCACGCTCGAAGACCT 114466
Qy 172 uileGlnAsnMetAsp-----ArgValAspArgGluIleThrMe 185
Db 114467 TCTGACGCGCTCAGCTCGATGATGAGCGGTGCGGCTCGGCTCCACCGCGCATCCGCC 114526
Qy 185 tValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeuGluGluAaL 205
Db 114527 CCGCGCGGTACCGGTCCCGG---TCCGCGCGGTCCGCGCGCTCCGCCG-TCTCTGCC 114582
Qy 205 aLysProProGluProGluLysProValSerPro----- 216
Db 114583 ATGTCGCGCTCTCTCAGCTCGGCTCCCTCAGCTCCCGTACTCCGTAGCAACGCGAGGCCG 114642
Qy 217 -ProProIleGluSerLysHisArgSerLeuValGlnIleIleTyraSpGluAsnArgLy 236
Db 114643 CCGCGCAGGCGCGCGCGCGCGCGCACA-----GAGCGCGCGG 114681
Qy 236 sLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLe 256
Db 114682 CCGCGAGCGCGCGGTGAACGCCACCCACCGCGGCGAGTCCAGCGGATACGCGCGCCCA 114741
Qy 256 uTyr-----AsnGlnProSerAspThrArgGln---TyrHisGluAsnIleLysI 272
Db 114742 GCACATAGCTCGCGCGGCGAGCGGTTACGAAACGCCGCGGCGAGCAACGTCGCCCGC 114801
Qy 272 eAsnGlnAlaMetArgLysLysLeuIleLeuTyraPheLysArgArgAsnHisAlaArgLy 292
Db 114802 GCACGAGTC-----GAGCGCGAACACGCGTCCGCG 114831
Qy 292 sGlnTrpLysGlnLys-----PheCysGlnArgTyraSpGlnLeuMetGluAlaLe 309
Db 114832 GGTACTGCAGCATCTGCTACCGCGCGTGAAGCGGTGGCCACCTCCGAGGCGTCT 114891
Qy 309 uGluLysLysValGluArgIleGluAsnAsnProArgArg-----ArgAlaLysG 326
Db 114892 GCGCGCGAACTGGACGC-----CCCGCGCGCACCAACACCGCGCAGAAGA 114939
Qy 326 uSerLysValArgGluTyTyTyGluLysGlnPheProGluIleArgLysGlnArgGluLe 346
Db 114940 TCCCGCGCGCGTGTATGCGCGCACCGGCGAGCAGCAGCACCTTCGCGCGC---CGTCCAGT 114996
Qy 346 uGlnGluArgMetGlnSerValGlyGlnArgGlySerGlyLeuSerMetSerAlaL 366
Db 114997 CGATGTGCGAGCACCGCGCGGTACCGGAGCACCGAGCCCTGCGACACCGCGCGCA 115056
Qy 366 aArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuG 386
Db 115057 GCGCGCGCGCGCGAA-----CCGCTCCCGCGCACCTGCGCGAGCACCGCGCGCGAC 115110
Qy 386 uLysGlnMetArg---GlnLeuAlaValIleProProMetLeuTyraSpGlnGln 405
Db 115111 GCACGAGCAGCTGTCTCAGGCTGCGGTGCGCGACCCG----- 115147
Qy 405 nArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyLysAs 425
Db 115148 -----GCGCGCGCGCGTTCAT---GGAGCGCA 115173
Qy 425 pArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMe 445
Db 115174 TCGCCAGTCT-----GGCCAGCGCGAACA----- 115198
Qy 445 tGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaG 465
Db 115199 -----CACCCC-----CGACACCGGTACAGGA 115221
Qy 465 uCysValLeuTyTyTyTyLeuThrLysLysAsnGluAsnTyTyTySerLeuValArg 485
Db 115222 A-----CG 115224
Qy 485 gSerTyArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 505
Db 115225 CCACCTCGGGCAGTAGTACCGCGCGCGCGTCCACCGCGGAGAACATCAGCAGGATCG 115284

```

```
QY 505 nGlnGlnGlnGlnPnMetProArgSerSerGlnGluGluLysAspGluLysGlu 525
Db 115285 CCACGAAGTCAG-----CGCGGTATCGCGAAGTTCGCCGAACCGCGTCAGCG 115332
QY 525 sGluLysGluAlaGluLysGluGluLysProGluValGluAsnAspLysGluAspLe 545
Db 115333 CGAACGACGTCGCGTACGC---CATCGTCGACCTGATCCATCCCGCGCATCAGCCGCT 115389
QY 545 uLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaValAl 565
Db 115390 ACGCGCGACCCCTCGACAC----- 115411
QY 565 aSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThrArg-S 585
Db 115412 -----GGCGCGCGCGTACCCCGCGCGCGCGCGCTCGCGCGCT 115452
QY 585 erMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuA 605
Db 115453 CGCGCGCTCCGCGCGCTCCGCG-----GCCCTCTGCCACCTCAG 115491
QY 605 laSerMetGluLeuAsnGluSerSerArg-----TrpThrGluGluMetG 621
Db 115492 CCACCTTGACACACCCCTCGGTGCGCGCGACTGACACGCGCGCGCGCGCGCGCGCG 115551
QY 621 luThrAlaLysLysGlyLeuLeuGluHisGlyArg---AsnTrpSerAlaIleAlaArgM 640
Db 115552 AGCGCCACCGCCCGAGGACGCTGGAACAGATACGTCCCGAGCGATCCCGCTCCCGCAGC 115611
QY 640 etValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgG 660
Db 115612 AGCAGCTCCCGCGACCTCGAGCAA-----CGACACCCACGCG 115650
QY 660 lAsnLeuAspGluIleLeuGlnGlnHisLysLysLysMetGluLysGluArgAsnAlaA 680
Db 115651 AGCGCGCGACCGCTCGCGAGACCGCGCGGAACAGCTTCAGCGGACGACGATCCCG 115710
QY 680 rgArgLysLysLysAlaProAlaAlaIaSerGluGluAlaAlaPheProValV 700
Db 115711 GAGCAGAACAGCCCGCGCACCA-----CGCATCTCGGTACCCCGCGTCCG 115758
QY 700 alGluAspGluGlu-MetGluAlaSerGly-----ValSerLysGlnGluGluMet 717
Db 115759 TCGAGCAGCAGACGCGGACGCGGACCGACAGTAGCGGCGGCGGAATCCACCACTC 115818
QY 718 ValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSer 737
Db 115819 GCCAACACCGCCACCGAGAAAGCGCCACACGAGACCTCCCGCGGCGCGCGCAC 115878
QY 738 GlyProAlaThrValAsnAsnSerAspThrGluSerIleProSer-ProHisThrG1 757
Db 115879 GGGNAGAACAGCGACCGAAGACGAGCGGACCAACCGCGCGCGCGCGCGCGCGCGCG 115938
QY 757 uAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspG1 777
Db 115939 GCCCGCC-----CCACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 115977
QY 777 yProProProGly-----ProProThrPr 785
Db 115978 TCCGCGGACGGTACAGATCGCGGCGACATCACCCTGACGGATGGCTCCATCAACTCG 116037
QY 785 oProArg-----ArgThrSerArgAlaProIleGluProThrProAlaSerG1 801
Db 116038 TCCTCGAACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 116097
QY 801 u-----AlaThrGlyAlaProThr----- 807
Db 116098 AACGTACCGCGCTGCGCTGCTATCTCTCCGAGCTGCGGCTTCTCTGTCGCCACAGGCG 116157
QY 807 ----- 807
Db 116158 AGATAGGTGTAAACAGGATCAAAACCGNAGACCGTATTGGTGAACACCCCGCGCGGGTG 116217
```

```
QY 808 -----ProProProAlaProProSerProSerAl 817
Db 116218 GCCGCGGATACGTGCGTACCGTCGGAATCCCGCGCGCGCGCGCGCGCGCGCGCG 116277
QY 817 aProProPro-----ValValProLysGluGluLysGluGluGluThrAla-AlaA 834
Db 116278 CCGAGCCCGCGCATCTCAGCCCTCCCGCTCCGACCGCGCACCAAGAGCGGAGGCTAGTG 116337
QY 834 laProProValGluGluGlyGluGlu-----GlnLysProp 846
Db 116338 CGCCCCCGGGGGCCCTGCGACGGAATTCACGGGTGTGCGGTGCGGTGGGAACGGAAC 116397
QY 846 roAlaAlaGluLeuAlaValAspThrGlyLysAlaGluGlu-ProValLysSerGlu 865
Db 116398 CGCGGTGCGAGAGTCTTCCACGAAGAACTCAGAGCTTCCGAAACCGCTACGAGCGAA 116457
QY 866 CysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAla 885
Db 116458 GCA-----GAAGCCGAAACAGAGGTCCGTGCACGAGATGAGCGAGCGGACGCGCA 116511
QY 886 ThrAlaGluGlyAlaLeuLysAlaGluLys-----LysGluGly 898
Db 116512 GCCCGCAACCGAGGTGGCGCCGAGAGAGCCCGAGCGCGCGCGCGCGCGCGCGCG 116571
QY 899 GlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSer 918
Db 116572 GGAACAGGCGAGCGCCCGGAACGCCGCGGAACCGCGGAAGCGGAGAGAGCGCGCGCG 116631
QY 919 -----AlaThrCysSerAlaAspGluValAspGluAlaGlu----- 930
Db 116632 CAAGCGCACCGGATGCGCGCATGTATCCGACCTGCGCATCGTCTGAGCACCGCTCGT 116691
QY 931 -----GlyLysAspLysAsnArgLeuLeuSerProArgProSerLeuLeuThrPro 947
Db 116692 CGTCGCGCTGCTGCTGATCGTGGCGGCTCTT---CTTCGG-----CTACACACT 116739
QY 948 ThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLysGlnLeuLys 967
Db 116740 CGTGCAGATCCCGCGCGCAACGCGCTCGCCACCAAGCAGTCCACGCTCTACCTGTACGC 116799
QY 968 GlnArgAlaAlaIleProProIleGlnValThrLysValHisGluProProArgGlu 987
Db 116800 CGACGCGCA-GCCTCTCTCGCC-----GCGACGCGGAGGTCAACCGCGAGAGCG 116846
QY 988 AspAlaAlaProThrLysProAlaProAlaProPro-----ProProGlnAsnLeu 1005
Db 116847 TCGGCTCTCCCGCATCTCCAAGGCGCGCGAGCGCGCTCTCTCGCGCGCGGAGCGCG 116906
QY 1006 GlnProGluSerAspAlaProGlnGlnPro-----GlySerSer 1018
Db 116907 ACTTCTACAGCGAGTCCGCGCATCGACCCCATGGCGATGTCGCGGCGCGCTGGAACACG 116966
QY 1019 ProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGlu 1038
Db 116967 CCACCGCGAAGGCGAGCAGTCCGCTCCACCATCA----- 117002
QY 1039 AlaGlnLysLeuProGlyAspProProCysTyrThrSerGly-LeuProPheProValPr 1058
Db 117003 -----CCGACGATGCTCAAGAACTACTACTACTGCGCGCGCGCGCG 117041
QY 1058 oProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAl 1078
Db 117042 AGACCGTCA-----CCGCAAGGTGAAGGAGTCTTCTCATCTCGATCAAGC 117086
QY 1078 aProProGlyHis-----ProLeuProLeuGlyLeuHisAspThrAlaAr 1093
Db 117087 TCGACCGGAGAGAGACGAGAGATCCTCGAGGCTACCTCAACACAGCTTCTTCG 117146
QY 1093 gProValLeuProArgProProThrIleSerAsnProProLeuIleSerSerAlaLy 1113
Db 117147 G-----CCGCGCGCG-CCTACGCGCATCCAGCGCGCGCGCGCGCTACTTCTCGCGCTC 117196
QY 1113 sHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLe 1133
```



```
Db 120601 CGTGTGACGAACAGGGCAAGTCTGAACGGCCCTGTACAACGTCCGGCCCA---CCGG 120657
Qy 2379 laAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyG 2399
Db 120658 CCAGGTGGCCA-----AGATCATCAAGGACTTGGGGATCTCAGGGCG 120699
Qy 2399 ly-----GlyGlyLysAlaLysValSerGlyArgPro----- 2409
Db 120700 CTGCCCCAGCAGTAGCATGCGGGGAGCAGCGCC-----GGCGGGCGGTGTGGA 120750
Qy 2410 -----SerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProp 2428
Db 120751 ATTGGCAGTCACGCTGGGTTTAGGTCCAGTGGGGTAACTCCCGTGAGGGTT-CGAGTCC 120809
Qy 2428 roSerValSerSerValHisSerGluGlyAspCysAsnArgAlaGlyThrProLeuThrAsnA 2448
Db 120810 CTCGGCGCCACCGTA-----CGTTTCCGATGCTCAGCC 120845
Qy 2448 rgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyAsnProLeuI 2468
Db 120846 GAGCAGCTCGCGCACCAACCGGACCGCCCGGACGCT----- 120886
Qy 2468 leMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAla---- 2486
Db 120887 -----TGCGCGGGTGGGTGATCGCGTTCTTCTCTCGCGCGTCAAGTTCCGCGCACGT 120938
Qy 2487 --GlySerGlyProLeuAlaGly 2493
Db 120939 TCGGCTCGCCCTCCGGCTGGA 120961
```

Search completed: April 16, 2004, 10:50:23
Job time : 42585 secs

This Page Blank (uspto)